2017

Nonparametric Inference for Orderings and Associations Between two Random Variables

Chuan-Fa Tang
University of South Carolina

Nonparametric inference for orderings and associations between two random variables

by

Chuan-Fa Tang

Bachelor of Science
National Taiwan University 2007

Master of Science
National Taiwan University 2010

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Statistics

College of Arts and Sciences

University of South Carolina

2017

Accepted by:

Joshua M. Tebbs, Major Professor

Dewei Wang, Major Professor

John Grego, Committee Member

Alexander McLain, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

Dedication

In memory of my father

Tang, Su-Hsiang

Acknowledgments

Thanks to my respectable mentors

Dr. Joshua Tebbs and Dr. Dewei Wang

for their tireless forging

Thanks to my committee members

Dr. John Grego and Dr. Alexander McLain

for their helpful suggestions

Thanks to my dear mother and brother

Wang, Chin-Hao and Tang, Ming-Siang

for their kind support

Thanks to my dear father

Tang, Su-Hsiang

for his selfless dedication

to keep my dreams alive and continuing

Abstract

Ordering and dependency are two aspects to describe the relationship between two random variables. In this thesis, we choose two hypothesis testing problems to tackle; i.e., a goodness-of-fit test for uniform stochastic ordering and one for positive quadrant dependence. For the test for uniform stochastic ordering, we propose new nonparametric tests based on ordinal dominance curves. We derive the limiting distributions of test statistics and provide the least favorable configuration to determine critical values. Numerical evidence is presented to support our theoretical results, and we apply our methods to a real data set. An extension for random right-censored data is provided. For the test for positive quadrant dependence, we propose empirical-likelihood-based testing approaches. Without the need to estimate or smooth distribution or copula functions, our proposed testing procedure is more straightforward than previous methods. Simulation results show that our proposed tests are competitive in realistic settings. Stock price data sets are provided for illustration. An extension to test for exchangeability is provided.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Literature review

For a bivariate random vector (X, Y), one is usually interested in questions of how X and Y are related. In this dissertation, we address two questions. If X and Y are independent, which one is larger? If they are not independent, how might we adequately describe the dependence structure?

One way to compare the marginal distributions of X and Y is to compare their means. A two sample t-test can be used to order the means under normal assumptions. However, without the normal distribution assumption, ordering the means may not be sufficient to describe the ordering between the distributions. To compare distributions more generally, we introduce three commonly used stochastic orderings in this dissertation.

Directly comparing the cumulative distribution functions of X and Y leads to one of the most intuitive stochastic orderings, called ordinary stochastic ordering. A dominating relationship between two marginals reveals that one random variable "tends to provide larger values" than the other. Two stronger stochastic orderings, uniform stochastic and likelihood ratio ordering, are introduced by comparing conditional distributions. For uniform stochastic ordering, conditional on both X and Y being larger than a fixed number, one random variable tends to provide larger values. This type of ordering is useful to determine if, for example, one medical treatment is "uniformly" better than another. For likelihood ratio ordering, conditional on both X

and Y being between two fixed numbers, one random variable tends to provide larger values. This type of ordering is useful theoretically and has important applications in finance and econometrics. It can be shown easily that likelihood ratio ordering is the strongest ordering among these three and that ordinary stochastic ordering is the weakest.

From the last paragraph, we can see that ease of interpretation is one of the advantages of thinking in terms of stochastic orderings. More than that, order-restricted estimators are well-developed for each, and most of these estimators are better than their unrestricted versions when the corresponding ordering holds. For example, under ordinary stochastic ordering, El Barmi and McKeague (2005) developed restricted estimators for distribution functions and Davidov and Herman (2012) developed restricted estimators for the area under the ordinal dominance curve (Bamber, 1975). They also showed that their estimators have smaller mean squared errors than those that are unrestricted. For uniform stochastic ordering restriction, Rojo and Samaniego (1993), Mukerjee (1996), Rojo (2004), and El Barmi and Mukerjee (2016) proposed restricted estimators of distribution functions. Rojo (2004) and El Barmi and McKeague (2005) showed that their proposed estimators have smaller mean squared errors.

Before using a restricted estimator, it is crucial to know if the corresponding stochastic ordering holds; otherwise, this may introduce unnecessary bias. In this dissertation, we focus on uniform stochastic ordering as described in Chapter 2. To reveal the existence of this ordering, we consider goodness-of-fit hypothesis problems; i.e., testing the uniform stochastic ordering assumption versus not. Previously proposed testing procedures (e.g., Park et al., 1998; Arcones and Samaniego, 2000) lose power due to discretizing the supports or overestimating critical values. To gain more power, we propose a new way to tackle this hypothesis testing problem.

We use the ordinal dominance curve to describe stochastic orderings and adapt the majorant-based testing approaches described in Carolan and Tebbs (2005), Davidov and Herman (2012) and Beare and Moon (2015). In Chapter 2, we find that majorant-based methods for uniform stochastic ordering inherit good properties. Test statistics are easy to compute, and most importantly, the least favorable configurations exist (which helps to determine critical values and control Type I error probability). We provide an extension to incorporate random right-censored data in Chapter 4.

In this dissertation, we also focus on the positive quadrant dependence structure proposed by Lehmann (1966). A bivariate random vector (X, Y) is positive quadrant dependent (PQD) if the probability of X and Y of being simultaneously small is at least as large as it would be when X and Y are independent. The PQD property has useful applications in finance, insurance, and risk management. For example, if the true dependence structure of (X, Y) is PQD, then insurance premiums involving the portfolio containing X and Y will be underestimated if X and Y are treated as independent. See Dhaene and Goovaerts (1996), Denuit et al. (2001), Denuit and Scaille (2004), Embrechts et al. (2002) for more discussion. For applications in reliability theory and other ares, see Levy (1992), Shaked and Shanthikumar (1994), Drouet and Kotz (2001), and Lai (2003).

By unifying a joint distribution function, a copula function condenses the dependence structure of X and Y within the unit square $[0, 1]^2$. To determine the existence of PQD, it therefore suffices to consider this function. Most previous work estimates a copula and the departure from the null hypotheses. Janic-Wròbelwska et al. (2004) parametrized this structure and performed tests for PQD. To test against PQD, Denuit and Scaille (2004) and Scaillet (2005) proposed testing procedures with different functional distances. Gijbels et al. (2010) further used the estimators proposed by Omelka et al. (2009) to test against PQD. Gijbels and Sznajder (2013) developed a new resampling method with a PQD restriction to determine proper critical values. However, estimating the copula or the departure from a null hypothesis could be problematic if kernel estimation is involved, as it depends on bandwidth selection and is therefore potentially time-consuming and subjective.

To avoid these problems, we develop new hypothesis tests using empirical likelihood (Owen, 1990) to determine the existence of PQD between X and Y. Based on the empirical likelihood approach, many testing problems have been studied and have been shown to be powerful (Einmahl and McKeague, 2003; El Barmi and McKeague, 2013). We also expect these testing procedures to be powerful when testing for and against PQD.

1.2 Outline

In Chapter 2, we propose nonparametric goodness-of-fit tests for uniformly stochastic ordering with two continuous distributions based on the L^p difference between the sample ordinal dominance curve and its least star-shaped majorant. We then derive asymptotic distributions and prove that our testing procedure has a unique least favorable configuration for $p \in [1, \infty]$. In Chapter 3, we focus on hypothesis testing problems for PQD. We use empirical likelihood and derive the asymptotic distributions of test statistics, while suggesting limiting least favorable configurations to construct rejection regions. Preliminary numerical results and a comparison with Gijbels and Sznajder (2013) are provided. In Chapter 4, we consider an extension to test against uniform stochastic ordering with censored data and propose a new test for exchangeability. Lemmas and critical values are given in appendices.

Chapter 2

Nonparametric goodness-of-fit tests for uniform stochastic ordering

Summary: We propose L^p distance-based goodness-of-fit (GOF) tests for uniform stochastic ordering with two continuous distributions F and G, both of which are unknown. Our tests are motivated by the fact that when F and G are uniformly stochastically ordered, the ordinal dominance curve $R = FG^{-1}$ is star-shaped. We derive asymptotic distributions and prove that our testing procedure has a unique least favorable configuration of F and G for $p \in [1, \infty]$. We use simulation to assess finite-sample performance and demonstrate that a modified, one-sample version of our procedure (e.g., with G known) is more powerful than the one-sample GOF test suggested by Arcones and Samaniego (2000, *Annals of Statistics*). We also discuss sample size determination. We illustrate our methods using data from a pharmacology study evaluating the effects of administering caffeine to prematurely born infants.

2.1 Introduction

Suppose X and Y are continuous random variables with distribution functions F and G, respectively. In many applications, it is of interest to compare F and G. The ordinal dominance curve (ODC), which plots $(G(t), F(t))$ for $-\infty \leq t \leq \infty$, is a useful graphical tool that facilitates such a comparison (Bamber, 1975; Hsieh and Turnbull, 1996; Carolan and Tebbs, 2005; Davidov and Herman, 2012). The ODC can also be defined as $R = FG^{-1}$, where $G^{-1}(u) = \inf\{t : G(t) \geq u\}$ is the quantile function of G. When $F = G$, the ODC follows the main diagonal of the unit square, the so-called equal distribution line.

We consider order-restricted comparisons of F and G. Define $\overline{F} = 1 - F$ and $\overline{G} = 1 - G$. These are the survivor functions if X and Y are lifetime random variables, although herein we do not require X and Y to be nonnegative. Denote the corresponding densities by f and g, respectively. If $\overline{F} \leq \overline{G}$, then X and Y are stochastically ordered; this is written as $F \leq_{\mathrm{S}} G$ and means informally that X "tends to be smaller" than Y. Two stronger orders are the uniform stochastic order and the likelihood ratio order. When $\overline{F}/\overline{G}$ is nonincreasing, X and Y satisfy a uniform stochastic order, written $F \leq_{\mathrm{US}} G$. When f/g is nonincreasing, X and Y satisfy a likelihood ratio order, written $F \leq_{\mathrm{LR}} G$. It is easy to show these orderings follow the nested structure: $F \leq_{\mathrm{LR}} G \Longrightarrow F \leq_{\mathrm{US}} G \Longrightarrow F \leq_{\mathrm{S}} G$. A comprehensive account of these and other orderings is given in Shaked and Shanthikumar (2007).

Different stochastic orderings give rise to different functional forms of the ODC. The weakest ordering $F \leq_{\mathrm{S}} G$ holds if and only if R is at least as large as the equal distribution line; i.e., $R(u) \geq u$, for $0 \leq u \leq 1$. The strongest ordering $F \leq_{\mathrm{LR}} G$ holds if and only if R is concave. The intermediate ordering $F \leq_{\mathrm{US}} G$ holds if and only if R is *star-shaped* (Lehmann and Rojo, 1992). One way to characterize a star-shaped ODC is that the slope of the secant line from the point $(1, 1)$ to $(u, R(u))$; i.e., $r(u) = \{1 - R(u)\}/(1 - u)$, is nonincreasing in u. Figure 2.1 gives examples of ODCs that correspond to stochastic, uniform stochastic, and likelihood ratio orderings. This figure demonstrates the utility of the ODC in characterizing how two distributions are ordered and how the structure $F \leq_{\mathrm{LR}} G \Longrightarrow F \leq_{\mathrm{US}} G \Longrightarrow F \leq_{\mathrm{S}} G$ manifests itself graphically in the ODC.

This chapter is motivated by a pharmacology study evaluating the effects of administering caffeine to prematurely born infants in Columbia, South Carolina; see Section 2.5. Among 404 infants in the study, $m = 127$ were administered caffeine

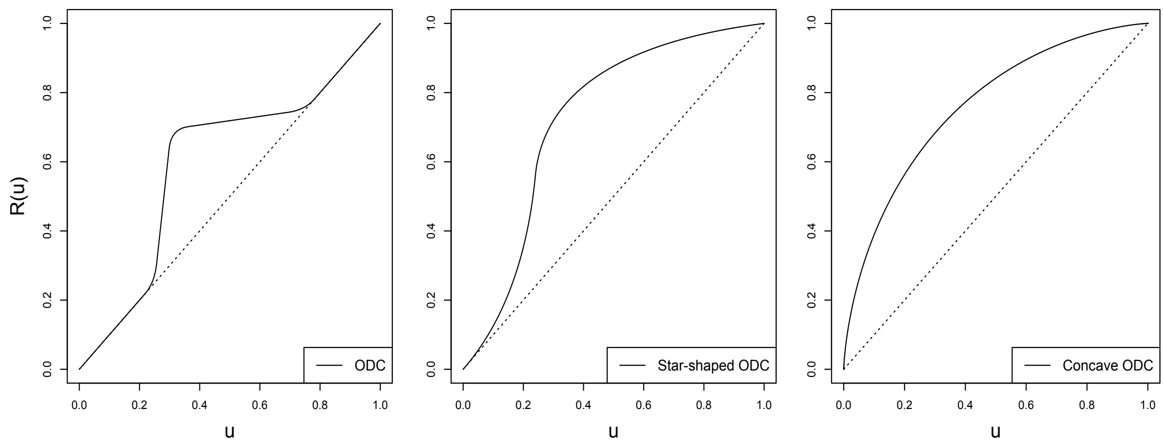

Figure 2.1: Ordinal dominance curves. Left: $F \leq_S G$. Middle: $F \leq_{US} G$. Right: $F \leq_{LR} G$. In each subfigure, the equal distribution line is shown dotted.

and $n = 277$ were not. Each infant was then followed until he or she was discharged from the hospital. All infants were eventually discharged and were alive at the time of discharge; i.e., no discharge times were censored. One of the goals of the study was to understand how the distributions of discharge times F (caffeine) and G (no caffeine) compared for the two groups. In Figure 2.2 (left), we display the sample ODC for the data, which is defined as $R_{mn}(u) = F_m\{G_n^{-1}(u)\}$, for $0 \leq u \leq 1$, where F_m and G_n are the empirical distribution functions and $G_n^{-1}(u) = \inf\{t : G_n(t) \geq u\}$ is the empirical quantile function. The sample ODC and its large-sample properties were described in Hsieh and Turnbull (1996).

On the basis of Figure 2.2, which stochastic ordering, if any, characterizes the true relationship between the discharge time distributions? There is a substantive literature on nonparametric tests for stochastic orderings with two or more distributions; see Davidov and Herman (2012), El Barmi and McKeague (2005), and the references therein. In the two-sample case, most of this literature describes tests where the equal distribution assumption $F = G$ is treated as the null hypothesis and the ordering (i.e., $F \leq_{\text{S}} G$, $F \leq_{\text{US}} G$, or $F \leq_{\text{LR}} G$) is placed in the alternative. A potential drawback with this type of test is that it is constructed assuming a specific

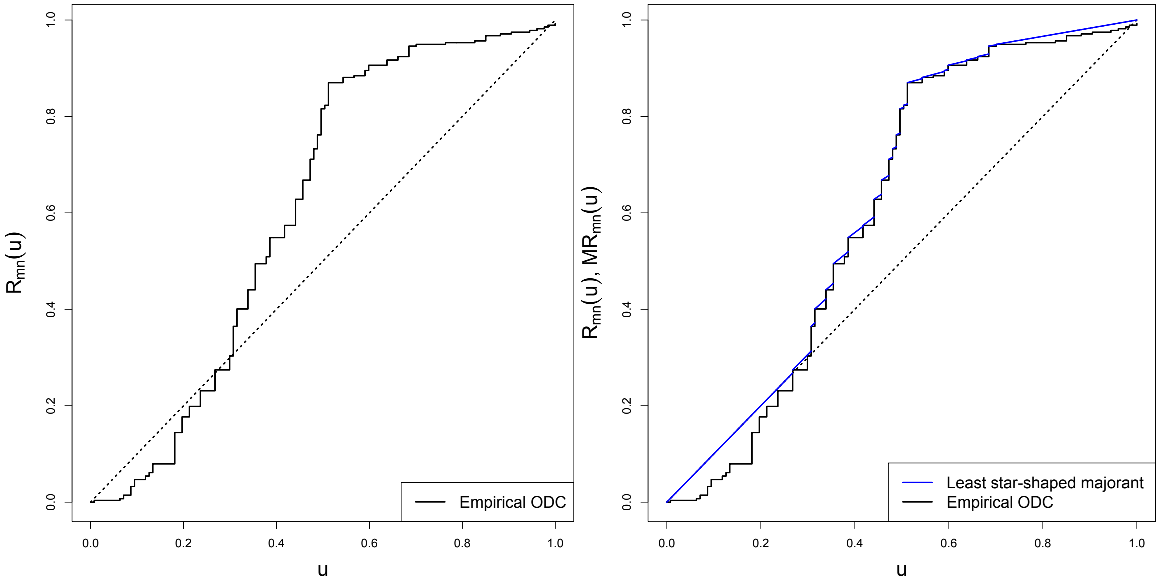

Figure 2.2: Premature infant data. Left: The sample ODC $R_{mn}(u) = F_m\{G_n^{-1}(u)\}$ for the time to discharge (F = caffeine; G = no caffeine). Right: The least star-shaped majorant $\mathcal{M}R_{mn}$ is shown in blue. In each subfigure, the equal distribution line is shown dotted.

order-restricted class of alternatives; if the assumed class is incorrect, the test may lead to misleading or vacuous conclusions. For example, applying tests of this type to the premature infant data, we obtain the following results:

- testing $F = G$ versus $F \leq_{\mathrm{S}} G$: p-value < 0.00002 (Davidov and Herman, 2012)
- testing $F = G$ versus $F \leq_{\mathrm{US}} G$: p-value < 0.00001 (Arcones and Samaniego, 2000)
- testing $F = G$ versus $F \leq_{\mathrm{LR}} G$: p-value < 0.00001 (Carolan and Tebbs, 2005).

Each test clearly dictates that the infant data are not consistent with $F = G$. However, we are no closer to identifying which specific ordering (if any) holds in this setting.

In this light, we consider goodness-of-fit (GOF) testing procedures instead. By "goodness-of-fit," we mean the procedure places the ordering in the null hypothesis and attempts to detect departures from the ordering. By comparison, the literature on nonparametric GOF tests with two distributions is more sparse, perhaps because this type of testing problem is more difficult. The primary reason for the added difficulty is that the ordering can hold under different configurations of F and G. Therefore, one must determine the least favorable configuration of the two distributions before the test can be performed; i.e., so that the probability of type I error can be controlled. Carolan and Tebbs (2005) proposed nonparametric GOF tests for likelihood ratio ordering with two continuous distributions by using the least concave majorant of the sample ODC. This work was generalized and improved upon by Beare and Moon (2015) in the econometrics literature, who considered likelihood ratio ordering and its applications in finance.

GOF tests for uniform stochastic ordering have been proposed but only in limited settings. Dardanoni and Forcina (1998) considered likelihood-based tests against uniform stochastic ordering in a two-way contingency table. Park et al. (1998) used a nonparametric maximum likelihood approach to formulate GOF tests with two or more continuous distributions, but only after data from these distributions have been assigned to disjoint intervals in the form of counts. This essentially discretizes the problem and results in testing against uniform stochastic ordering among several multinomial distributions. Furthermore, this formulation gives rise to non-unique least favorable configurations that depend on how the intervals are selected, the number of distributions, and even the significance level used. Finally, in the two-population setting, Arcones and Samaniego (2000) suggested a GOF test for uniform stochastic ordering based on the family of order-restricted estimators in Mukerjee (1996). However, these authors assume that one of the population distributions is known (e.g., G is known) and do not determine the least favorable configuration for their procedure. Instead, the authors use critical values from an upper bound asymptotic distribution which leads to a conservative test.

In this chapter, we propose a family of GOF tests for uniform stochastic ordering with two continuous distributions F and G; that is, we are interested in testing $H_0 : F \leq_{\mathrm{US}} G$ versus $H_1 : F \not\leq_{\mathrm{US}} G$, where both distributions are unknown. Motivated by the ODC approaches taken in Carolan and Tebbs (2005) and Beare and Moon (2015), we construct test statistics for H_0 versus H_1 based on the L^p difference between the sample ODC and its *least star-shaped majorant* (defined in Section 2.2). We then derive asymptotic distributions and prove that our testing procedure has a unique least favorable configuration for $p \in [1, \infty]$. Interestingly, this theoretical result is different from the finding in Beare and Moon (2015), who showed that when using L^p distance-based GOF tests for likelihood ratio ordering, the least favorable configuration exists only when $p \in [1, 2]$. Furthermore, unlike Park et al. (1998), our approach does not require one to discretize the support of the distributions which can only lead to a loss in power. Finally, we show that the one-sample version of our test (e.g., with G known) is not as conservative as the test proposed by Arcones and Samaniego (2000) and is generally better equipped to detect departures from H_0.

Formulating L^p distance-based GOF tests for uniform stochastic ordering in the two-sample problem is technically challenging. It is not possible to simply modify the proofs in Carolan and Tebbs (2005) and Beare and Moon (2015) under likelihood ratio ordering; see Section 2.3. At the same time, establishing that such an ordering exists has great practical implications. For example, if X and Y are lifetime random variables (and are absolutely continuous), then $F \leq_{\mathrm{US}} G$ is equivalent to the corresponding hazard rates being ordered. This is an important characterization in reliability and survival analysis applications. Our interest in uniform stochastic ordering is motivated by our collaboration with researchers in the premature infant study discussed earlier. Letting X and Y denote the times to discharge for the caffeine and no-caffeine groups, respectively, uniform stochastic ordering holds if and only if $\mathrm{pr}(X > t | X > t_0) \leq \mathrm{pr}(Y > t | Y > t_0)$, for all t, t_0 satisfying $t > t_0 \geq 0$. In other words, no matter how much time $t_0 \geq 0$ has subsequently passed, administering caffeine is consistent with shorter discharge times. Note that, in this context, stochastic ordering requires that the relationship above hold only initially (i.e., when $t_0 = 0$). Uniform stochastic ordering guarantees this type of dominance will hold for all $t_0 \geq 0$.

2.2 Testing procedure

Suppose that $X_1, X_2, ..., X_m$ are independent and identically distributed (iid) from F and that $Y_1, Y_2, ..., Y_n$ are iid from G. We assume the two samples are independent and that both F and G are unknown. Let $R = FG^{-1}$ denote the corresponding ODC. For our asymptotic results in Section 2.3 to hold, as in Hsieh and Turnbull (1996), we assume F and G have continuous densities f and g and that the first derivative of R is bounded over $[0, 1]$. Throughout this chapter, we denote the parameter space of R by Θ, the collection of nondecreasing, continuously differentiable functions from $[0, 1]$ to $[0, 1]$. Under our assumptions, the hypotheses $H_0 : F \leq_{\text{US}} G$ and $H_1 : F \not\leq_{\text{US}} G$ can be expressed equivalently as

$$H_0 : R \in \Theta_0 = \{\theta \in \Theta : \theta \text{ is star-shaped}\} \quad \text{and} \quad H_1 : R \in \Theta_1 = \Theta \setminus \Theta_0.$$

Recall that $\theta \in \Theta$ is star-shaped if and only if $\{1 - \theta(u)\}/(1 - u)$ is nonincreasing in u.

Let $R_{mn} = R_{mn}(u) = F_m\{G_n^{-1}(u)\}$ denote the sample ODC, defined in Section 2.1. Informally, our testing procedure is based on measuring the distance between R_{mn} and an estimate of R subject to the constraint that $F \leq_{\text{US}} G$. Towards defining this restricted estimator, let $l([0, 1])$ denote the collection of bounded functions on $[0, 1]$. For any $h \in l([0, 1])$, its least star-shaped majorant is defined as

$$\mathcal{M}h = \inf\{h^* \in l([0, 1]) : h \leq h^* \text{ and } h^* \text{ is star-shaped}\};$$

i.e., $\mathcal{M}h$ is the smallest star-shaped function in $l([0, 1])$ that is at least as large as h. Throughout our work, we call $\mathcal{M} : l([0, 1]) \mapsto l([0, 1])$ the least star-shaped majorant operator. Just as R_{mn} is an estimator of R under no restriction (Hsieh and Turnbull, 1996), the least star-shaped majorant $\mathcal{M}R_{mn}$ is an estimator of R under $H_0 : F \leq_{\text{US}} G$. Using Lemma A.1 in Appendix A, we show that this restricted estimator can be calculated as

$$\mathcal{M}R_{mn}(u) = 1 - \min_{\substack{v \in \mathcal{V} \cup \{0\} \\ v \leq u}} \left\{ \frac{1 - R_{mn}(v)}{1 - v} \right\} (1 - u),$$

for $0 \leq u < 1$, where $\mathcal{V}$ is the set of discontinuous (jump) points of R_{mn} and $\mathcal{M}R_{mn}(1) = 1$. Figure 2.2 (right) shows the least star-shaped majorant of the sample ODC for the premature infant data described in Section 2.1.

Our testing procedure utilizes the sample ODC R_{mn} and its least star-shaped majorant $\mathcal{M}R_{mn}$. Specifically, we propose the family of test statistics

$$M_{mn}^p = c_{mn} \|\mathcal{M}R_{mn} - R_{mn}\|_p,$$

where $c_{mn} = \{mn/(m+n)\}^{1/2}$ is a normalizing constant and $\|\cdot\|_p$ is the L^p norm with respect to Lebesgue measure. We allow for $p \in [1, \infty]$; i.e., $\|h\|_p = (\int_{[0,1]} |h(u)|^p du)^{1/p}$ when $p < \infty$ and $\|h\|_\infty = \sup_{u \in [0,1]} |h(u)|$. For example, when $p = 1$, $\|\mathcal{M}R_{mn} - R_{mn}\|_1$ equals the area between the two estimators; when $p = \infty$, $\|\mathcal{M}R_{mn} - R_{mn}\|_\infty$ equals the largest vertical distance between the estimators. For any $p \in [1, \infty]$, clearly large values of M_{mn}^p are evidence against H_0.

2.3 Theoretical results

In this section, we first describe the asymptotic distribution of M_{mn}^p for any star-shaped ODC; i.e., for any $R \in \Theta_0$. We then demonstrate that, for any $p \in [1, \infty]$, all null distributions are dominated stochastically by the asymptotic distribution of M_{mn}^p under $R(u) = u$, that is, when $F = G$. From this least favorable distribution, we can find the critical value $c_{\alpha,p}$ that satisfies $\lim_{m,n\to\infty} \text{pr}(M_{mn}^p \geq c_{\alpha,p}) = \alpha$ when $F = G$ and $\lim_{m,n\to\infty} \text{pr}(M_{mn}^p \geq c_{\alpha,p}) \leq \alpha$ when $H_0 : F \leq_{\text{US}} G$ is true. In other words, rejecting H_0 when $M^p_{mn} \geq c_{\alpha,p}$ is an asymptotic size α decision rule. Finally, we examine relevant asymptotic distributions when $R \in \Theta_1$ and then characterize large-sample power properties. We also discuss sample size calculations to detect departures from H_0. All theorems are proved in Section 2.7. Additional technical details are provided in Appendix A.

2.3.1 Asymptotic results under H_0

Let $\mathcal{I}$ denote the identity operator on $l([0,1])$ and define $\mathcal{D} = \mathcal{M} - \mathcal{I}$. When H_0 is true; i.e., when $R \in \Theta_0$, note that $\mathcal{M}R = R$ and

$$M^p_{mn} = c_{mn}\|\mathcal{M}R_{mn} - R_{mn}\|_p = c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p.$$

At first glance, establishing the limiting distribution of M^p_{mn} under H_0 might seem to be straightforward, that is, one could simply start with the asymptotic distribution of $c_{mn}(R_{mn} - R)$ described in Hsieh and Turnbull (1996) and apply the functional delta method (see, e.g., Section 2.3.9 in Vaart and Wellner, 1996) and continuous mapping theorem. This was the approach taken by Beare and Moon (2015) with their L^p distance-based GOF test statistics under likelihood ratio ordering. In our setting, this direct approach is not possible because whereas the least concave majorant operator in Beare and Moon (2015) is Hadamard directionally differentiable (Shapiro, 1990; Shapiro, 1991), the least star-shaped majorant operator $\mathcal{M}$ (and hence $\mathcal{D}$) is not always so; see Lemma A.5 in Appendix A. Fortunately, this does not create insurmountable problems because weak convergence of $c_{mn}(\mathcal{D}R_{mn} - \mathcal{D}R)$ is not a necessary prerequisite to derive the asymptotic distribution of $c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p$.

Before we state the asymptotic distribution of M^p_{mn} for any $R \in \Theta_0$, we need to describe R precisely because these distributions depend completely on the shape of R. Recall that when $R \in \Theta_0$, the slope function $r(u) = \{1 - R(u)\}/(1 - u)$ is nonincreasing in u. When $r(u)$ is strictly decreasing over $[0,1]$, we say that R is *strictly* star-shaped. When $R \in \Theta_0$ is not strictly star-shaped, then, analogously to Beare and Moon (2015), there exists a unique collection (finite or countable) of closed, pairwise disjoint intervals of the form $[a_k, b_k]$, $0 \leq a_k < b_k \leq 1$, where

- the slope $r(u)$ is constant over each interval (i.e., R is affine over each interval)
- no two intervals possess the same value of $r(u)$.

In this case, we say that $R \in \Theta_0$ is *non-strictly* star-shaped. The reason we bifurcate Θ_0 using "strictly" and "non-strictly" descriptors is that the nondegenerate part of the asymptotic distribution of M^p_{mn} depends only on those regions where R is non-strictly star-shaped. If R is strictly star-shaped over $[0,1]$, the distribution of M^p_{mn} collapses to zero in the limit.

To make our description of the asymptotic distributions precise, we therefore introduce the following notation. For $0 \leq a < b \leq 1$, define

$$\mathcal{M}^{(1,0)}_{[a,b]} h = \inf\{h^* \in l([0,1]) : h \leq h^* \text{ and } h^* \text{ is star-shaped} \\ \text{over } [a,b] \text{ with kernel } (1,0)\}.$$

A general definition of what it means for a function h^* to be star-shaped with kernel (c,d) is given directly before Lemma A.1 in Appendix A. For any $h \in l([0,1])$, the function $\mathcal{M}^{(1,0)}_{[a,b]} h$ has two defining characteristics. First, $\mathcal{M}^{(1,0)}_{[a,b]} h(u) = h(u)$ whenever $u \notin [a,b]$. Second, over $[a,b]$, $\mathcal{M}^{(1,0)}_{[a,b]} h$ is the smallest function (at least as large as h) that is star-shaped with kernel $(1,0)$; i.e., the slope function $-\mathcal{M}^{(1,0)}_{[a,b]} h(u)/(1-u)$ over $[a,b]$ is nonincreasing in u. The importance of the functional operator $\mathcal{M}^{(1,0)}_{[a,b]} : l([0,1]) \mapsto l([0,1])$ becomes clear as we state our first main result.

Theorem 2.1. *Suppose $R \in \Theta_0$ and let $\mathcal{B}$ denote a standard Brownian bridge. The asymptotic results below hold when* $\min\{m,n\} \to \infty$ *and* $n/(m+n) \to \lambda \in (0,1)$.
(a) *If R is strictly star-shaped over* $[0,1]$, *then* $M^p_{mn} \xrightarrow{d} 0$ *for all* $p \in [1,\infty]$.
(b) *If R is non-strictly star-shaped, then for* $p \in [1,\infty)$,

$$M^p_{mn} \xrightarrow{d} \left\{ \sum_k \left[\lambda R'(a_k) + (1-\lambda)\{R'(a_k)\}^2\right]^{p/2} \int_{a_k}^{b_k} \left\{\mathcal{D}^{(1,0)}_{[a_k,b_k]} \mathcal{B}(u)\right\}^p du \right\}^{1/p};$$

when $p = \infty$,

$$M_{mn}^p \xrightarrow{d} \sup_k \left\{ \left[\lambda R'(a_k) + (1-\lambda)\{R'(a_k)\}^2\right]^{1/2} \sup_{u \in [a_k, b_k]} \left\{ \mathcal{D}_{[a_k,b_k]}^{(1,0)} \mathcal{B}(u) \right\} \right\}.$$

In both asymptotic distributions, R' is the derivative of R and $\mathcal{D}_{[a_k,b_k]}^{(1,0)} = \mathcal{M}_{[a_k,b_k]}^{(1,0)} - \mathcal{I}$.

From Theorem 2.1, one can see that when $F \leq_{\mathrm{US}} G$, the only randomness in the asymptotic distribution of M_{mn}^p arises from the non-strictly star-shaped regions $[a_k, b_k]$ and is described probabilistically by the $\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}$ processes. Furthermore, when $F = G$, the asymptotic distribution of M_{mn}^p simplifies to $\|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_p$ for all $p \in [1, \infty]$. When $p = 1$, for example, this quantity describes the distribution of the area between the least star-shaped majorant of a standard Brownian bridge $\mathcal{B}$ and $\mathcal{B}$ itself. When $p = \infty$, $\|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_\infty$ describes the distribution of the sup-norm distance between these two processes. Readers familiar with the GOF tests for likelihood ratio ordering in Carolan and Tebbs (2005) and Beare and Moon (2015) will no doubt recognize the homology between our Theorem 2.1 and the corresponding results in these articles. However, as noted earlier, GOF tests for uniform stochastic ordering present their own set of mathematical challenges and different conclusions are reached about the existence of a least favorable configuration.

Theorem 2.2. *Suppose $R \in \Theta_0$. For any $p \in [1, \infty]$, the asymptotic distribution of M_{mn}^p is ordinary stochastically smaller than $\|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_p$; i.e.,*

$$\lim_{\substack{m,n\to\infty \\ n/(m+n)\to\lambda}} \mathrm{pr}_{R\in\Theta_0}(M_{mn}^p \geq t) \leq \mathrm{pr}(\|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_p \geq t),$$

for all $t \in \mathbb{R}$, where λ is defined in Theorem 1.

Theorem 2.2 establishes that when using M_{mn}^p to test $H_0 : F \leq_{\mathrm{US}} G$ versus $H_1 : F \not\leq_{\mathrm{US}} G$, the equal distribution line $R(u) = u$ represents the least favorable configuration of F and G for all $p \in [1, \infty]$. Proving this result involves showing that each of the $\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}$ processes in Theorem 2.1 are mutually independent, a somewhat startling discovery because each process shares the same Brownian bridge $\mathcal{B}$ and each operator $\mathcal{D}^{(1,0)}_{[a_k,b_k]}$ shares the same kernel point $(1,0)$. The practical utility of Theorem 2.2 is that, for any $p \in [1,\infty]$, we can determine the critical value that maximizes the probability of type I error over all configurations of F and G in Θ_0. This result is different than the conclusion reached in Beare and Moon (2015), who showed that when testing against likelihood ratio ordering using L^p distance-based statistics involving the least concave majorant of R_{mn}, $R(u) = u$ is the least favorable configuration when $p \in [1, 2]$ and for $p > 2$ the least favorable configuration does not exist. Careful inspection of Theorem 2.1 and some intuition sheds insight on why this is true. When R is star-shaped, but not strictly star-shaped, each of the derivatives $R'(a_k)$ in Theorem 2.1 satisfies $R'(a_k) \leq 1$. However, when $F \leq_{\mathrm{LR}} G$, there is no guarantee these derivatives are uniformly bounded for all concave R and hence anomalous limiting behavior can result when p is too large.

For given values of the significance level α and $p \in [1,\infty]$, denote the $1-\alpha$ quantile of $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$ by $c_{\alpha,p}$; i.e., $c_{\alpha,p}$ solves $\alpha = \mathrm{pr}(\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p \geq c_{\alpha,p})$. To approximate the distribution of $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$, we generated 100,000 Brownian bridge paths on a grid of 100,000 equally spaced points in $[0, 1]$, and, for each $p \in \{1, 2, 3, 5, \infty\}$, we calculated $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$ for each path. For each p, these 100,000 values were used to approximate the density function of $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$ and quantiles $c_{\alpha,p}$, for $\alpha = 0.01$, 0.05, and 0.10. These functions and the selected quantiles $c_{\alpha,p}$ are provided in Appendix A.

2.3.2 Asymptotic results under H_1

The difference between the asymptotic distribution of M^p_{mn} under $H_0 : R \in \Theta_0$ and that under $H_1 : R \in \Theta_1$ arises from the non-star-shaped regions of R. To characterize a non-star-shaped ODC $R \in \Theta_1$, start with $\mathcal{M}R$, which is star-shaped, and note that (as in Section 2.3.1) one can partition the unit interval $[0, 1]$ as $[0, 1] = S \cup (\cup_k S_k)$, where $\mathcal{M}R$ is strictly star-shaped over S and non-strictly star-shaped over pairwise disjoint intervals of the form $S_k = [a_k, b_k]$, $0 \le a_k < b_k \le 1$, for $k = 1, 2, \ldots.$ One can further partition each S_k as $S_k = S_{k1} \cup S_{k2}$, where $S_{k1} = \{u \in S_k : \mathcal{M}R(u) = R(u)\}$ and $S_{k2} = \{u \in S_k : \mathcal{M}R(u) > R(u)\}$. Each S_{k1} must contain a_k so it is never empty, and the non-star-shaped regions of R can be written as $\cup_k S_{k2}$. In other words, $R \in \Theta_0$ when $\cup_k S_{k2}$ is empty and $R \in \Theta_1$ otherwise.

In general, these types of regions contribute differently to the limiting distribution of M^p_{mn}. Over the strictly star-shaped region S, $\mathcal{M}R(u) = R(u)$ for all u and the L^p norm of $c_{mn}\{\mathcal{D}R_{mn}(u) - \mathcal{D}R(u)\}$ converges in distribution to 0, as in Section 2.3.1. To clearly describe the contribution over the S_k regions, we introduce new notation. For any $h \in l([0,1])$, define the functional operator $\mathcal{L}_{S_k} : l([0,1]) \mapsto l([0,1])$ according to

$$\mathcal{L}_{S_k} h(u) = -\inf_{\substack{v \in S_{k1} \\ v \le u}} \left\{ \frac{-h(v)}{1-v} \right\} (1-u) I_{S_k}(u) + h(u) I_{S_k^c}(u), \quad \text{for } u \in [0,1),$$

where $I_A(\cdot)$ is the indicator function over the set A and A^c denotes the complement of A. When $u = 1$, $\mathcal{L}_{S_k} h(u) = \max\{h(1), 0\}$ or $h(1)$ depending on whether the singleton $\{1\} \in S_{k1}$ or not; see Appendix A. Using this new operator, we now characterize asymptotic distributions for any ODC $R \in \Theta$ with those in $\Theta_1 = \Theta \setminus \Theta_0$ of particular interest. A discussion on the large-sample power properties of our testing procedure follows.

Theorem 2.3. *Suppose $R \in \Theta$. Using the notation described in this subsection,*

$$c_{mn} \|\mathcal{D}R_{mn} - \mathcal{D}R\|_p \xrightarrow{d} \left\{ \sum_k \int_{u \in S_k} \left| \mathcal{L}_{S_k} T_R^\lambda(u) - T_R^\lambda(u) \right|^p du \right\}^{1/p}$$

for $p \in [1, \infty)$; when $p = \infty$,

$$c_{mn} \|\mathcal{D}R_{mn} - \mathcal{D}R\|_p \xrightarrow{d} \sup_k \sup_{u \in S_k} \left| \mathcal{L}_{S_k} T_R^\lambda(u) - T_R^\lambda(u) \right|.$$

Both results hold as $\min\{m,n\} \to \infty$ and $n/(m+n) \to \lambda \in (0,1)$. In both cases, $T_R^\lambda(u) = \lambda^{1/2}\mathcal{B}_1(R(u)) + (1-\lambda)^{1/2} R'(u) \mathcal{B}_2(u)$, $0 \le u \le 1$, where $\mathcal{B}_1$ and $\mathcal{B}_2$ denote two independent standard Brownian bridges.

Four remarks are in order. First, the process $T_R^\lambda = \{T_R^\lambda(u), 0 \leq u \leq 1\}$ in Theorem 2.3 is well known; as noted earlier, it represents the asymptotic distribution of $c_{mn}(R_{mn} - R)$ for any $R \in \Theta$; see, e.g., Theorem 2.2 in Hsieh and Turnbull (1996). Second, the asymptotic distributions identified in Theorem 2.3 apply for any $R \in \Theta$, but we show in Appendix A that they quickly reduce to those in Theorem 2.1 when $R \in \Theta_0$. Third, our L^p tests are consistent for $p \in [1, \infty]$. To see why, consider the sup-norm ($p = \infty$) case in Theorem 2.3 and note that, by the triangle inequality,

$$\begin{aligned}\mathrm{pr}_{R\in\Theta_1}(M_{mn}^\infty \geq c_{\alpha,\infty}) &= \mathrm{pr}_{R\in\Theta_1}\Big(c_{mn}\left\|\mathcal{D}R_{mn}\right\|_\infty \geq c_{\alpha,\infty}\Big)\\ &\geq \ \mathrm{pr}_{R\in\Theta_1}\Big(c_{mn}\left\|\mathcal{D}R_{mn} - \mathcal{D}R\right\|_\infty \leq c_{mn}\left\|\mathcal{D}R\right\|_\infty - c_{\alpha,\infty}\Big)\end{aligned}$$

which can be approximated by

$$\mathrm{pr}_{R\in\Theta_1}\Bigg(\sup_k \sup_{u\in S_k}\Big|\mathcal{L}_{S_k}T_R^\lambda(u) - T_R^\lambda(u)\Big| \leq c_{mn}\left\|\mathcal{D}R\right\|_\infty - c_{\alpha,\infty}\Bigg).$$

It is easy to show that $\sup_k \sup_{u\in S_k}|\mathcal{L}_{S_k}T_R^\lambda(u) - T_R^\lambda(u)|$ is bounded and that, for any $R \in \Theta_1$, $c_{mn}\|\mathcal{D}R\|_\infty \to \infty$, as $\min\{m, n\} \to \infty$, which establishes our claim. The finite p argument is analogous. Fourth, approximate lower bounds on the power, like the one above in the sup-norm case, can be used for sample size calculations. For an ODC $R \in \Theta_1$ deemed to be clinically relevant, one can determine numerically the smallest m and n that solve $\mathrm{pr}_{R\in\Theta_1}(\sup_k \sup_{u\in S_k}|\mathcal{L}_{S_k}T_R^\lambda(u) - T_R^\lambda(u)| \leq c_{mn}\|\mathcal{D}R\|_\infty - c_{\alpha,\infty}) = 1-\beta$, where $\beta \in (0, 1)$. The resulting solution will be inexorably conservative but still potentially useful for planning purposes. We illustrate this approach with examples in Section 2.4.

We conclude this section with a brief discussion on local power. This discussion is ultimately not dissimilar from the local power discussion in Beare and Moon (2015) under likelihood ratio ordering. However, our interest in local power arises because we want to compare the one-sample version of our testing procedure to the GOF test suggested by Arcones and Samaniego (2000). This one-sample comparison is given in Section 2.4.3. The two-sample discussion is given now. Let $\{R^{(r)}, r = 1, 2, ..., \}$ denote a sequence of ODCs in Θ_1. For each $r \geq 1$, denote the corresponding distributions by $F^{(r)}$ and $G^{(r)}$ from which we have independent random samples $X_1^{(r)}, X_2^{(r)}, ..., X_m^{(r)}$ and $Y_1^{(r)}, Y_2^{(r)}, ..., Y_n^{(r)}$, respectively. We examine local power properties by letting $R^{(r)}$ approach Θ_0 in the sense that $\|\mathcal{D}R^{(r)}\|_p = \|\mathcal{M}R^{(r)} - R^{(r)}\|_p \to 0$ as $r \to \infty$ at different rates. Using the notation in this paragraph, our last theorem summarizes the salient results.

Theorem 2.4. *Suppose the first derivative of $R^{(r)} \in \Theta_1$ is uniformly bounded over $[0, 1]$ for all r. Suppose $p \in [1, \infty]$. All limits stated below assume that $\max\{m, n\} = O(r)$ and $n/(m + n) \to \lambda \in (0, 1)$, as $r \to \infty$.*

(a) *If $\lim c_{mn}\|\mathcal{D}R^{(r)}\|_p = \infty$, then $\lim \mathrm{pr}_{R^{(r)}\in\Theta_1}(M^p_{mn} > c_{\alpha,p}) = 1$.*

(b) *For any $\beta \in (0, 1)$, there exists $\eta_p(\beta) > 0$ such that*

$$\liminf \mathrm{pr}_{R^{(r)}\in\Theta_1}(M^p_{mn} > c_{\alpha,p}) \geq 1 - \beta$$

whenever $\liminf c_{mn}\|\mathcal{D}R^{(r)}\|_p \geq \eta_p(\beta)$.

Part (a) of Theorem 2.4 indicates that when $\|\mathcal{D}R^{(r)}\|_p$ converges to 0 at a rate slower than c_{mn}^{-1}, $c_{mn}\|\mathcal{D}R^{(r)}\|_p$ diverges and the power of our test converges to 1. Part (b) guarantees that when $c_{mn}\|\mathcal{D}R^{(r)}\|_p$ remains bounded away from zero, the power of our test is still nontrivial; i.e., it does not converge to 0. This occurs when the "amount of information" c_{mn} increases and the "departure" $\|\mathcal{D}R^{(r)}\|_p$ decreases, and both do so at the same rate.

2.4 Simulation evidence

We use simulation to assess the finite-sample performance of our tests. In Section 2.4.1, we consider fixed ODCs under both $H_0 : F \leq_{\mathrm{US}} G$ and $H_1 : F \not\leq_{\mathrm{US}} G$ to estimate type I error probability and power, respectively, and we illustrate the sample size calculations described in Section 2.3.2. In Section 2.4.2, we modify our testing

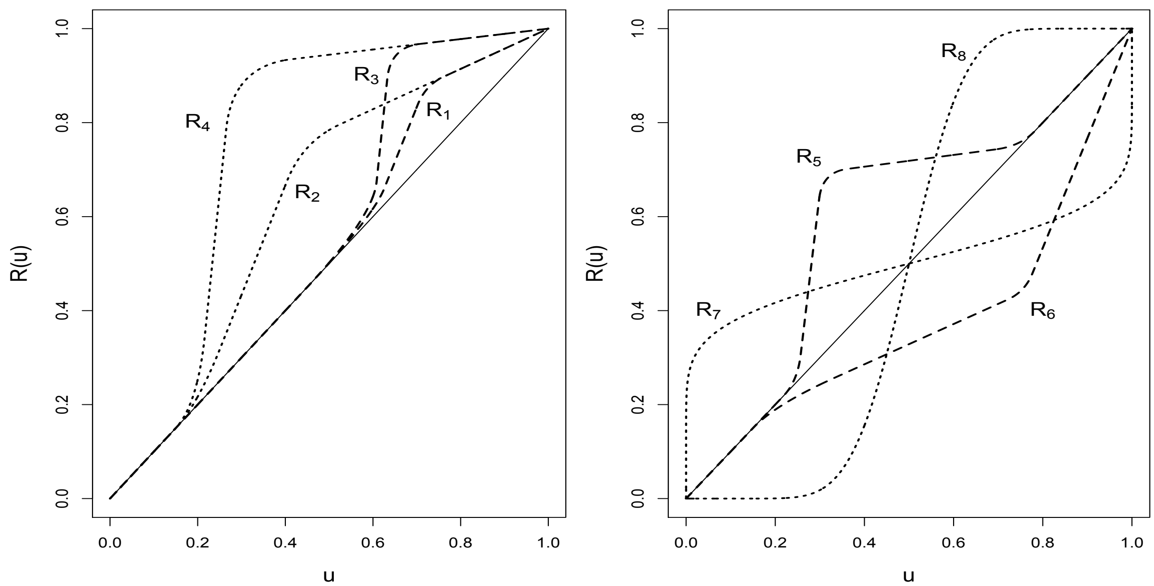

Figure 2.3: Left: Star-shaped ODCs; i.e., $R_i \in \Theta_0$. Right: Non-star-shaped ODCs; i.e., $R_i \in \Theta_1$. A description of each curve is given in Appendix A.

procedure to allow for one of the population distributions to be known and compare this modified test to the one-sample GOF test in Arcones and Samaniego (2000). Local power results are provided in Section 2.4.3.

2.4.1 Fixed ODC comparisons

We consider four ODCs satisfying $R \in \Theta_0$ (R_1, R_2, R_3, and R_4) and four ODCs satisfying $R \in \Theta_1$ (R_5, R_6, R_7, and R_8). The H_0 ODCs (Figure 2.3, left) are each members of a family of star-shaped ODCs that we describe in the Appendix A. The H_1 ODCs (Figure 2.3, right) are not star-shaped and are also described in Appendix A. We also consider $R_0 = R_0(u) = u$, for $u \in [0, 1]$, to examine finite-sample performance under the least favorable configuration $F = G$. All of our results are based on $10,000$ Monte Carlo data sets using independent samples from F and G with sample sizes m and n, respectively. To generate the samples, we let $F(u) = R_i(u)$ and $G(u) = u$, for $u \in [0, 1]$. We then sample $X_1, X_2, ..., X_m$ from F using the inverse cumulative distribution function technique and independently sample $Y_1, Y_2, ..., Y_n$

from a uniform$(0,1)$ distribution. This provides independent samples for each ODC R under consideration.

Table A.2 in Appendix A gives Monte Carlo estimates of the probability of rejecting $H_0 : F \leq_{\mathrm{US}} G$ for different sample sizes, values of $p \in \{1, 2, \infty\}$, and $\alpha = 0.05$. We experimented with other values of p (i.e., $p = 3$ and $p = 5$) but obtained results similar to those when $p = 2$. Of initial interest is the finite-sample performance when $F = G$. With 10,000 simulated data sets, the margin of error associated with the size estimates under $F = G$, assuming a 99 percent confidence level, is approximately 0.006. Therefore, one notes that our tests with $p = 1$ and $p = 2$ are slightly anti-conservative with small samples and otherwise operate closely to the nominal level. Furthermore, examining the rejection rates for the other star-shaped ODCs (R_1, R_2, R_3, and R_4) supports Theorem 2.2 which, for $p \in [1, \infty]$, guarantees the probability of type I error will be at its maximum under $F = G$. Likewise, powers for the non-star-shaped ODCs (R_5, R_6, R_7, and R_8) all approach unity as m and n become large. This reinforces our consistency claim.

We also use the non-star-shaped ODCs in Figure 2.3 to illustrate sample size determination. For $p \in [1, \infty]$ and for a given $R \in \Theta_1$, denote by $d_{R,\beta,p}$ the $1 - \beta$ quantile of the asymptotic distributions in Theorem 2.3. Using our lower bound on the asymptotic power from Section 2.3.2 and taking $m = n$ (for simplicity), we obtain a closed-form expression for the minimum sample size necessary to detect the departure $\|\mathcal{D}R\|_p = \|\mathcal{M}R - R\|_p$ with probability $1 - \beta$ when using an asymptotic size α test; i.e.,

$$m = 2\left(\frac{d_{R,\beta,p} + c_{\alpha,p}}{\|\mathcal{D}R\|_p}\right)^2, \quad \text{for } p \in [1, \infty].$$

With $\alpha = 0.05$ and $1 - \beta = 0.8$, Appendix A tables these solutions for each non-star-shaped ODC in Figure 2.3 and for each $p \in \{1, 2, \infty\}$. For example, for the R_5 ODC, which corresponds to F and G being stochastically ordered (but not uniformly stochastically ordered), the minimum sample size solutions for $p \in \{1, 2, \infty\}$, respectively, are $m = 634$, $m = 461$, and $m = 582$. Such sample sizes might seem dispiritingly large; however, it is not surprising these solutions are conservative. We describe in Section 2.6 alternative approaches that should reduce this conservatism.

2.4.2 Comparison with Arcones and Samaniego (2000)

We now turn our attention to the special case of testing $H_0 : F \leq_{\mathrm{US}} G$ versus $H_1 : F \not\leq_{\mathrm{US}} G$ where G is known. Arcones and Samaniego (2000), who focused largely on optimal estimation of F (with $F \leq_{\mathrm{US}} G$ and G known), also suggested a conservative large-sample procedure to test against H_0. Their proposed test statistic, which we denote by D_m, can be expressed as a function of the one-sample ODC $R_m = F_m G^{-1}$; specifically,

$$D_m = m^{1/2} \sup_{0 \leq v \leq u \leq 1} [(1-v)\{1 - R_m(u)\} - (1-u)\{1 - R_m(v)\}].$$

However, instead of deriving a least favorable (asymptotic) distribution for inference, the authors proved that the asymptotic distribution of D_m is bounded above by $2 \sup_{u \in [0,1]} |\mathcal{B}(u)|$, where $\mathcal{B}$ is a standard Brownian bridge, and selected their critical value $c^{\mathrm{AS}}_{\alpha/2}$ to satisfy $\alpha = \mathrm{pr}(\sup_{u \in [0,1]} |\mathcal{B}(u)| \geq c^{\mathrm{AS}}_{\alpha/2})$. On the other hand, one-sample versions of our GOF procedure are available and use the test statistics

$$M^p_m = m^{1/2} \left[\int_{[0,1]} \{\mathcal{D} R_m(u)\}^p du \right]^{1/p} \quad \text{and} \quad M^\infty_m = m^{1/2} \sup_{u \in [0,1]} \{\mathcal{D} R_m(u)\},$$

where $\mathcal{D}$ is the operator defined in Section 2.3.1 and $R_m(u) = F_m\{G^{-1}(u)\}$. The limiting distributions in Theorem 2.1 also apply here as $m \to \infty$; in addition, it is straightforward to modify the proof of Theorem 2.2 to conclude that $F = G$ admits the least favorable configuration for $p \in [1, \infty]$ in the known G case.

For different sample sizes m (now corresponding to F only), Table A.3 in Appendix A gives small-sample rejection rates of our one-sample tests and the test from Arcones and Samaniego (2000), both performed using $\alpha = 0.05$. We used techniques similar to those described in Section 2.3.1 to approximate the critical value $c^{\mathrm{AS}}_{\alpha/2} = c^{\mathrm{AS}}_{0.025} = 1.359$

and performed all simulations in the same way as before except G is now known. Clearly, there is a price to be paid for using the test based on the D_m statistic when $F = G$; type I error probability estimates remain significantly below the nominal level for all $m \leq 200$. On the other hand, our $p = 1$ and $p = 2$ tests are only minimally conservative when $m \leq 75$, and our sup-norm ($p = \infty$) test performs nominally even when $m = 20$. In addition, the sup-norm test can be markedly more powerful at detecting non-star-shaped alternatives with small to moderately sized samples.

2.4.3 Local power analysis

A consequence of Theorem 2.3 is that, for any fixed $R \in \Theta_1$, our L^p GOF tests are consistent for all $p \in [1, \infty]$. To glean additional insight on which values of p might be preferred in practice, we investigate the power associated with local alternatives. Starting in the lower left corner, Figure 2.4 depicts a sequence of ODCs in Θ_1 that approach Θ_0 (moving from lower left to upper right). Each ODC shown in Figure 2.4 belongs to a family of ODCs described in Appendix A; the defining feature of this family is that it is indexed by a single parameter $\delta \in [0, 0.5]$. The $\delta = 0$ member, say $R_{(0)}$, is the initial ODC in the lower left corner of Figure 2.4; the $\delta = 0.5$ member $R_{(0.5)}$, shown in the upper right, is the limiting ODC in Θ_0. ODCs $R_{(\delta)}$ with intermediate values of $\delta \in (0, 0.5)$ are also identified in Figure 2.4.

In our testing problem, a local power analysis involves examining a sequence of ODCs $\{R^{(r)}, r = 1, 2, ..., \}$ in Θ_1 that converges to Θ_0 at different rates. We do so here by using the family of ODCs just described. Specifically, we consider the rates $\zeta_r \in \{\log r, r^{2/5}, r^{1/2}\}$. For each ζ_r, we first choose a sequence of constants $\delta^{(r)}$ such that $\lim_{r \to \infty} \zeta_r |\delta^{(r)} - 0.5| = c_{\zeta_r} > 0$ and then select members from our ODC family identified by $R^{(r)} = R_{(\delta^{(r)})}$, for $r = 1, 2, ...$ The resulting sequence $R^{(r)}$ satisfies $\|\mathcal{D}R^{(r)}\|_p = \|\mathcal{M}R^{(r)} - R^{(r)}\|_p \to 0$ and $\zeta_r \|\mathcal{D}R^{(r)}\|_p \to c^*_{\zeta_r, p} > 0$, both as $r \to \infty$. This investigation allows us to learn more about the practical aspects of Theorem 2.4

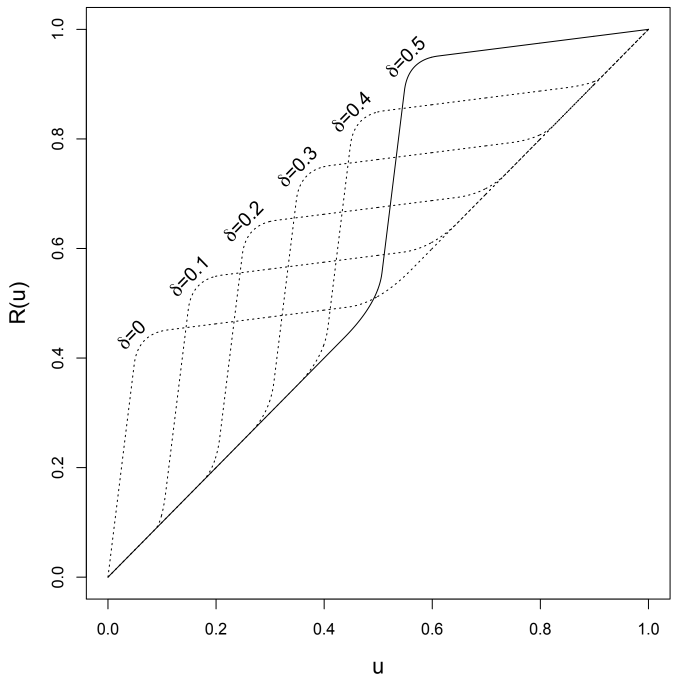

Figure 2.4: Local power family of ODCs indexed by $\delta \in [0, 0.5]$. The $\delta = 0$ member $R_{(0)}$ is the initial ODC in Θ_1; the $\delta = 0.5$ member $R_{(0.5)}$ is the limiting ODC in Θ_0. This family is described in Appendix A.

(i.e., with both F and G unknown). We also use these ODC sequences, one for each rate ζ_r, to compare the one-sample versions of our tests with the test in Arcones and Samaniego (2000).

For each $r \in \{50, 100, 500, 1000, 5000, 10000\}$, we simulated 10,000 independent random samples, $X_1^{(r)}$, $X_2^{(r)}$, ..., $X_m^{(r)}$ from $F^{(r)}$ and $Y_1^{(r)}$, $Y_2^{(r)}$, ..., $Y_n^{(r)}$ from $G^{(r)}$, where $F^{(r)}(u) = R^{(r)}(u)$ and $G^{(r)}(u) = u$, $0 \leq u \leq 1$, and $m = n = r$. Figure 2.5 (top row) shows the estimated powers of our $\alpha = 0.05$ tests associated with each rate: $\zeta_r = \log r$ (left), $\zeta_r = r^{2/5}$ (middle), and $\zeta_r = r^{1/2}$ (right). Note that with $m = n = r$, considering the slower rates $\zeta_r = \log r$ and $\zeta_r = r^{2/5}$ allows us to assess part (a) of Theorem 2.4, while the fastest rate $\zeta_r = r^{1/2}$ allows us to assess part (b). Both parts are supported by our empirical results in Figure 2.5. For the slower rates, the powers approach unity as expected; however, we find that there is no decisively preferred value of p among $p \in \{1, 2, \infty\}$. On the other hand, when $\zeta_r = r^{1/2}$, the $p = 1$

powers hover only slightly above 0.3 for all r, while the $p = 2$ and $p = \infty$ powers still approach unity.

Switching to the one-sample problem, we find quite different results. For each rate ζ_r, Figure 2.5 (bottom row) displays the estimated powers of our one-sample $\alpha = 0.05$ tests which use M_m^1, M_m^2, and M_m^∞. Powers were estimated in the same way as for the two-sample case except now we treat $G^{(r)}(u) = u$ as known and take $m = r$. In this setting, the sup-norm test consistently provides the largest power, followed by the $p = 2$ test and the $p = 1$ test. In addition, all three distance-based tests outperform the corresponding $\alpha = 0.05$ Arcones and Samaniego (2000) test in terms of local power, especially at the fastest rate $\zeta_r = r^{1/2}$ where $\mathrm{pr}_{R^{(r)} \in \Theta_1}(D_m > c_{0.025}^{\mathrm{AS}})$ appears to decrease towards zero.

2.5 Premature infant data

Caffeine is commonly used to treat newborn infants for apnea of prematurity (Schmidt et al., 2006) and to prevent the onset of respiratory distress syndrome, bronchopulmonary dysplasia, and extubation failure (Cox et al., 2015). Known as "the silver bullet" in the treatment of prematurely born infants at risk for these and other acute conditions (Aranda et al., 2010), caffeine is widely regarded within the neonatal care community to be safe and cost effective. It has also been approved by the United States Food and Drug Administration for use with preterm infants due to its history of providing beneficial outcomes with no long-term adverse side effects (Dobson and Hunt, 2013).

We now analyze the data from the study described in Section 2.1; for complete details, see Cox et al. (2015). Because assessing the use of caffeine with premature infants was a central focus of this study, we consider only those infants who were classified as "premature;" i.e., newborns whose gestational age was at or below 37 weeks. With F and G denoting the discharge time distributions for the caffeine and

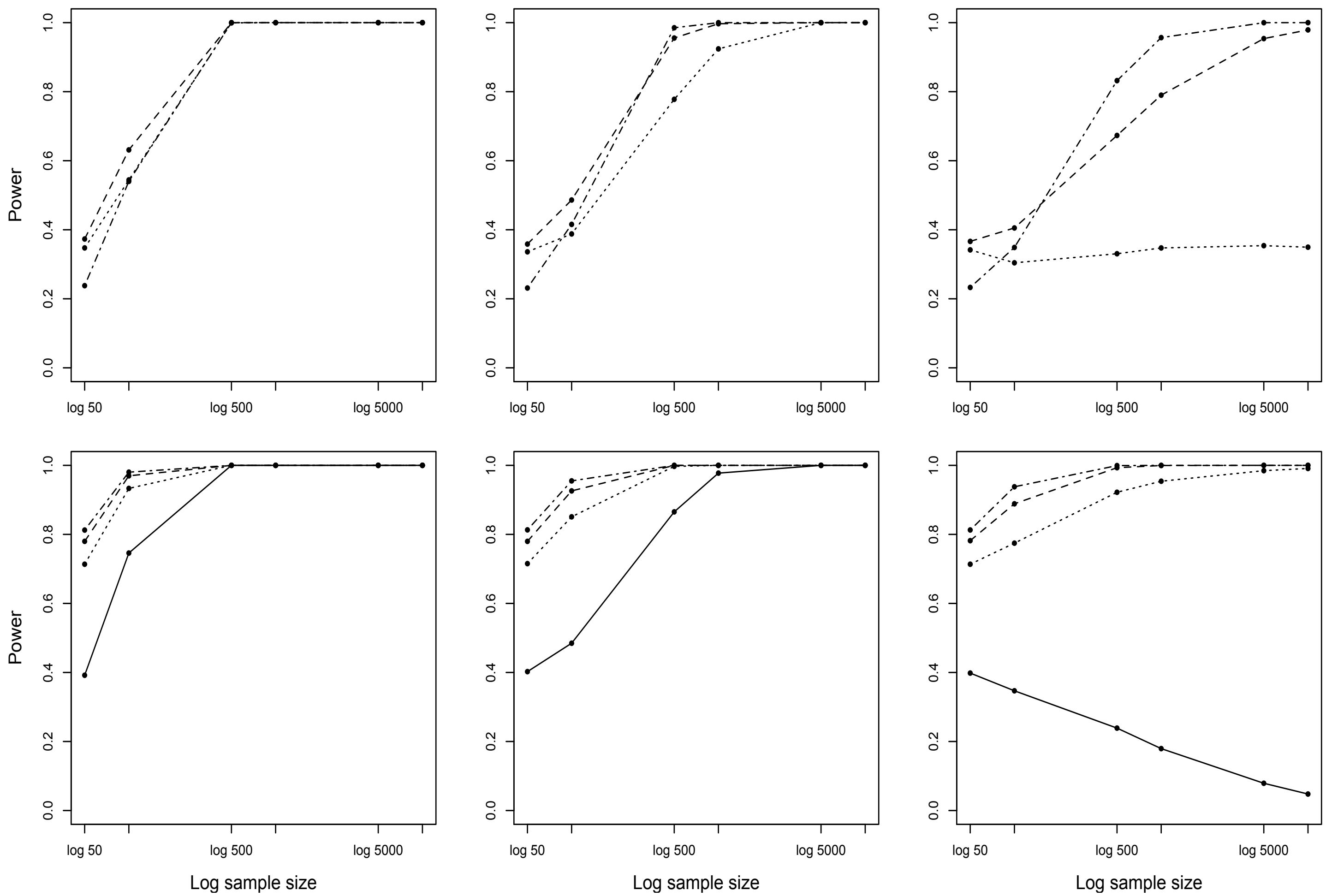

Figure 2.5: Local power results with $\alpha = 0.05$. Left: $\zeta_r = \log r$. Middle: $\zeta_r = r^{2/5}$. Right: $\zeta_r = r^{1/2}$. Top: Two-sample case. Bottom: One-sample case. Our L^p results are shown dotted for $p = 1$, dashed for $p = 2$, and dot-dashed for $p = \infty$. Arcones and Samaniego (2000) results (one-sample case only) are shown using a solid line.

no-caffeine groups, respectively, recall that Figure 2.2 displays the sample ODC R_{mn} and its least star-shaped majorant $\mathcal{M}R_{mn}$, calculated from samples of size $m = 127$ from F and $n = 277$ from G. As noted in Section 2.1, we performed the test in Davidov and Herman (2012) with these data and concluded that $F \leq_{\mathrm{S}} G$ was strongly supported over $F = G$. We also performed the GOF tests in Beare and Moon (2015) and concluded that $F \leq_{\mathrm{LR}} G$ would be rejected at $\alpha = 0.05$; the L^1 and L^2 statistics based on the least concave majorant of R_{mn} are 0.717 and 0.999, respectively, which are larger than the corresponding 0.95 quantiles 0.664 and 0.753 identified by their least favorable distributions.

We therefore assess whether or not the data in Figure 2.2 are consistent with uniform stochastic ordering. Testing $H_0 : F \leq_{\mathrm{US}} G$ versus $H_1 : F \nleq_{\mathrm{US}} G$ based on the least star-shaped majorant of R_{mn}, our GOF test statistics are $M^1_{mn} = 0.170$, $M^2_{mn} = 0.263$, and $M^\infty_{mn} = 0.949$, each of which is well below the $\alpha = 0.10$ critical values identified in Appendix A (0.496, 0.586, and 1.219, respectively), that is, H_0 cannot be discounted at any reasonable level of significance. Therefore, not only does caffeine therapy provide point-of-care health benefits and improved long-term outcomes for prematurely born infants, our analysis suggests that treating these infants with caffeine may also lead to hospital discharge times that are uniformly stochastically smaller than those for infants not treated with caffeine.

2.6 Concluding remarks

When two distributions F and G satisfy uniform stochastic ordering, F and G when conditioned on the interval $[t_0, \infty)$, for any $t_0 \in \mathbb{R}$, also satisfy uniform stochastic ordering. This desirable property could be exploited to increase the power of our tests under H_1 and simultaneously reduce the sample sizes necessary to detect departures from H_0. To see how, suppose that uniform stochastic ordering is suspected to be violated when $t > t_0$, either from historical information or from observing data in related applications. In this situation, one could apply our tests after conditioning to determine if R is non-star-shaped over the smaller region $[G^{-1}(t_0), 1]$ and calculate sample sizes to detect departures over it instead of over $[0, 1]$. A similar approach was suggested by Carolan and Tebbs (2005) for detecting departures from likelihood ratio ordering. In the same spirit, Beare and Moon (2015) suggest that bootstrapping samples over departure regions could help to increase the power of GOF tests for likelihood ratio ordering. This strategy may also be fruitful in our setting, allowing one to reduce the conservatism arising from relying on the least favorable distribution over the entire unit interval.

We believe that our GOF tests could be generalized to allow for different types of censored data, but the theory underpinning these extensions would not be trivial. For example, with random right-censored data, there would be nothing to prevent one from simply replacing the empirical survival functions $\overline{F}_m$ and $\overline{G}_n$ with Kaplan-Meier estimators of $\overline{F}$ and $\overline{G}$ and then calculating R_{mn} and $\mathcal{M}R_{mn}$ using these estimates. However, asymptotic distributions of the corresponding test statistics may depend heavily on the latent censoring distributions, and there is no guarantee that the least favorable configuration of F and G will exist. Future work could investigate censored-data extensions of majorant-based inference−not only with uniform stochastic ordering, but with other orderings as well.

Finally, estimating distributions under a uniform stochastic ordering assumption has received considerable attention for two populations; see, e.g., Rojo and Samaniego (1993), Mukerjee (1996), and Arcones and Samaniego (2000). We view the one- and two-sample tests proposed herein as helpful inference procedures to determine if the uniform stochastic ordering assumption is plausible and hence restricted estimation methods for F and G are warranted. An anonymous reader has suggested that developing pointwise confidence intervals for $R(u)$ under a uniform stochastic ordering constraint may be a worthwhile next step. We agree and comment on this further after

Lemma A.4 in Appendix A. Another interesting avenue for future research would be to generalize our majorant-based tests to more than two populations. Estimation techniques in this setting are available in Dykstra et al. (1991) and El Barmi and Mukerjee (2016).

2.7 Proofs

In this section, we provide the proofs of Theorems 2.1-2.4. Lemmas cited in this section are stated and proved in Appendix A.

Proof of Theorem 2.1. We start with the asymptotic distribution of R_{mn}, suitably centered and scaled. Applying Theorem 2.2 in Hsieh and Turnbull (1996), it follows that $c_{mn}(R_{mn} - R)$ converges weakly to T_R^λ as $\min\{m, n\} \to \infty$ and $n/(m+n) \to \lambda \in (0, 1)$, where T_R^λ satisfies $T_R^\lambda(u) = \lambda^{1/2}\mathcal{B}_1(R(u)) + (1-\lambda)^{1/2}R'(u)\mathcal{B}_2(u)$, for $0 \le u \le 1$, and $\mathcal{B}_1$ and $\mathcal{B}_2$ are independent standard Brownian bridges. When $R \in \Theta_0$, $\mathcal{D}R = 0$ and $M_{mn}^p = c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p$. Define the functional operator $d\mathcal{D}_R : l([0, 1]) \mapsto l([0, 1])$ by

$$d\mathcal{D}_R h(u) = \begin{cases} \max\{h(1), 0\} - h(1), & \text{if } u = 1 \\ \mathcal{M}_{[a_k, b_k]}^{(1,0)} h(u) - h(u), & \text{if } \exists k \text{ such that } a_k \le u \le b_k \\ 0, & \text{otherwise,} \end{cases}$$

for $h \in l([0, 1])$. Denote by $C([0, 1])$ the collection of all real continuous functions with domain $[0, 1]$. If $\mathcal{D}$ is Hadamard directionally differentiable tangentially to $C([0, 1])$ at R, then $d\mathcal{D}_R$ is the Hadamard directional derivative of $\mathcal{D}$. Applying the functional delta method and continuous mapping theorem yields $M_{mn}^p \xrightarrow{d} \|d\mathcal{D}_R T_R^\lambda\|_p$ for $p \in [1, \infty]$. Those situations in which $\mathcal{D}$ is Hadamard directionally differentiable are described in Lemma A.5 in Appendix A.

When $\mathcal{D}$ is not Hadamard directionally differentiable, the functional delta method and continuous mapping theorem cannot be applied. However, by using Lemma A.6 in

Appendix A, we are able to prove that $M^p_{mn} \xrightarrow{d} \|d\mathcal{D}_R T^\lambda_R\|_p$ anyway. For convenience, let $Z_{mn} = c_{mn}(R_{mn} - R)$ and $Z = T^\lambda_R$. From Theorem 12.2 in Billingsley (1999) and Skorohod's representation theorem (see, e.g., Theorem 6.7 in Billingsley, 1999), there exist random elements Z'_{mn} and Z' defined on a common probability space with $Z'_{mn} \overset{L}{=} Z_{mn}$ and $Z' \overset{L}{=} Z$ such that $\|Z'_{mn} - Z'\|_\infty \to 0$ almost surely. The notation "$\overset{L}{=}$" denotes that two processes are equivalent in distribution. Define $R'_{mn} = c^{-1}_{mn} Z'_{mn} + R$. From Lemma A.6 in Appendix A, because c^{-1}_{mn} decreases to 0 and $\|Z'_{mn} - Z'\|_\infty \to 0$ almost surely, then for all $p \in [1, \infty]$ we have

$$\lim_{\substack{m,n\to\infty \\ n/(m+n)\to\lambda}} c_{mn}\|\mathcal{D}R'_{mn} - \mathcal{D}R\|_p = \|d\mathcal{D}_R Z'\|_p$$

almost surely. Because $c_{mn}\|\mathcal{D}R'_{mn} - \mathcal{D}R\|_p \overset{d}{=} c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p$ and also $\|d\mathcal{D}_R Z'\|_p \overset{d}{=} \|d\mathcal{D}_R T^\lambda_R\|_p$, where the notation "$\overset{d}{=}$" means equal in distribution, we have

$$\lim_{\substack{m,n\to\infty \\ n/(m+n)\to\lambda}} c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p \overset{d}{=} \|d\mathcal{D}_R T^\lambda_R\|_p.$$

This shows that $M^p_{mn} \xrightarrow{d} \|d\mathcal{D}_R T^\lambda_R\|_p$ for all $p \in [1, \infty]$.

When R is strictly star-shaped over $[0, 1]$, it is easy to see that $\|d\mathcal{D}_R T^\lambda_R\|_p = 0$ which quickly establishes part (a). The remainder of the proof focuses on establishing part (b). When R is non-strictly star-shaped,

$$\|d\mathcal{D}_R T^\lambda_R\|_p = \left[\sum_k \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]} T^\lambda_R(u)\}^p du\right]^{1/p}$$

for $p \in [1, \infty)$ and $\|d\mathcal{D}_R T^\lambda_R\|_p = \sup_k\{\sup_{u\in[a_k,b_k]} \mathcal{D}^{(1,0)}_{[a_k,b_k]} T^\lambda_R(u)\}$ for $p = \infty$. Using Lemma A.1 in Appendix A, we write $\mathcal{D}^{(1,0)}_{[a_k,b_k]} T^\lambda_R(u) = \sup_{v\in[a_k,u]} Q_k(u, v)$ for $u \in [a_k, b_k]$, where

$$Q_k(u, v) = \left(\frac{1-u}{1-v}\right) T^\lambda_R(v) - T^\lambda_R(u), \quad \text{for } v \in [a_k, u].$$

In Lemma A.8 in Appendix A, we show that the processes

$$\{Q_k(u, v),\ a_k \le v \le u < b_k\}$$

are mutually independent across k. Therefore, $\{\mathcal{D}^{(1,0)}_{[a_k,b_k]}T^\lambda_R(u), u \in [a_k, b_k]\}$ are also mutually independent. To prove further results, we note that over each non-strictly star-shaped region $[a_k, b_k]$, we can write $R(u)$ as a linear function; i.e., $R(u) = 1 - R'(a_k)(1-u)$. Thus, from Lemma A.2 in Appendix A, we have

$$\mathcal{D}^{(1,0)}_{[a_k,b_k]}T^\lambda_R(u) = \mathcal{D}^{(1,0)}_{[a_k,b_k]}\{W^\lambda_R(u) - l^\lambda_{R,k}(1)\},$$

for all k, where $W^\lambda_R(u) = \lambda^{1/2}\mathcal{W}_1(R(u)) + (1-\lambda)^{1/2}R'(u)\mathcal{W}_2(u)$,

$$l^\lambda_{R,k}(u) = \lambda^{1/2}\{1 - R'(a_k)(1-u)\}\mathcal{W}_1(1) + (1-\lambda)^{1/2}R'(a_k)u\mathcal{W}_2(1),$$

and $\mathcal{W}_1$ and $\mathcal{W}_2$ are independent standard Wiener processes; i.e., $\mathcal{W}_i$, for $i = 1, 2$, satisfies $\mathcal{B}_i(u) = \mathcal{W}_i(u) - u\mathcal{W}_i(1)$, $0 \le u \le 1$, for $i = 1, 2$. Based on the properties of a standard Wiener process, it follows that for $u \in [a_k, b_k]$,

$$\begin{aligned}
\mathcal{W}_i(R(u)) - \mathcal{W}_i(1) &= \mathcal{W}_i(1 - R'(a_k)(1-u)) - \mathcal{W}_i(1) \\
&\overset{L}{=} R'(a_k)^{1/2}\{\mathcal{W}_i(u) - \mathcal{W}_1(1)\},
\end{aligned}$$

for $i = 1, 2$. Furthermore, for $u \in [a_k, b_k]$, we have $R'(u) = R'(a_k)$ and

$$\begin{aligned}
W^\lambda_R(u) - l^\lambda_{R,k}(1) &\overset{L}{=} \lambda^{1/2}R'(a_k)^{1/2}\{\mathcal{W}_1(u) - \mathcal{W}_1(1)\} \\
&\qquad\qquad + (1-\lambda)^{1/2}R'(a_k)\{\mathcal{W}_2(u) - \mathcal{W}_2(1)\} \\
&\overset{L}{=} \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{1/2}\{\mathcal{W}(u) - \mathcal{W}(1)\},
\end{aligned}$$

where $\mathcal{W}$ is a standard Wiener process. The last equivalence (in distribution) follows because both right-hand side processes above are Gaussian, they have the same mean $E\{W^\lambda_R(u) - l^\lambda_{R,k}(1)\} = 0$, for $u \in [a_k, b_k]$, and they have the same covariance $\mathrm{cov}\{W^\lambda_R(u_1) - l^\lambda_{R,k}(1), W^\lambda_R(u_2) - l^\lambda_{R,k}(1)\} = \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}\min\{1-u_1, 1-u_2\}$, for $u_1, u_2 \in [a_k, b_k]$. Using Lemma A.2 in Appendix A again, we have

$$\begin{aligned}
&\mathcal{D}^{(1,0)}_{[a_k,b_k]}\{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{1/2}\{\mathcal{W}(u) - \mathcal{W}(1)\} \\
&\qquad\qquad = \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{1/2}\mathcal{D}^{(1,0)}_{[a_k,b_k]}\mathcal{B}(u),
\end{aligned}$$

where $\mathcal{B}$ is a standard Brownian bridge formed by $\mathcal{W}$; i.e., $\mathcal{B}(u) = \mathcal{W}(u) - u\mathcal{W}(1)$, for $u \in [0,1]$. We can therefore write

$$\int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]} T^{\lambda}_R(u)\}^p du \ \stackrel{d}{=} \ \{\lambda R'(a_k) \ + \ (1 \ - \ \lambda) R'(a_k)^2\}^{p/2} \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]} \mathcal{B}(u)\}^p du,$$

for $p \in [1,\infty)$, and

$$\sup_{u\in[a_k,b_k]} \mathcal{D}^{(1,0)}_{[a_k,b_k]} T^{\lambda}_R(u) \stackrel{d}{=} \{\lambda R'(a_k) + (1-\lambda) R'(a_k)^2\}^{1/2} \sup_{u\in[a_k,b_k]} \mathcal{D}^{(1,0)}_{[a_k,b_k]} \mathcal{B}(u),$$

for $p = \infty$. For $p \in [1,\infty)$, we have shown that $\int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]} T^{\lambda}_R(u)\}^p du$ are mutually independent. One can show that $\int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]} \mathcal{B}(u)\}^p du$ are also mutually independent by replacing $T^{\lambda}_R(\cdot)$ with $\mathcal{B}(\cdot)$ in the definition of $Q_k(u,v)$ and repeating the same argument. Therefore, we have

$$\sum_k \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]} T^{\lambda}_R(u)\}^p du \stackrel{d}{=} \sum_k \{\lambda R'(a_k) + (1-\lambda) R'(a_k)^2\}^{p/2} \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]} \mathcal{B}(u)\}^p du,$$

which completes the proof for $p \in [1,\infty)$. Completing the proof for the $p = \infty$ case is analogous.

$\square$

Proof of Theorem 2.2. When $F = G$, the ODC is $R_0 = R_0(u) = u$, $0 \le u \le 1$, and $T^{\lambda}_{R_0} \stackrel{d}{=} \mathcal{B}$. Because R_0 is non-strictly star-shaped over $[0,1]$, Theorem 2.1 yields $M^p_{mn} \xrightarrow{d} \|d\mathcal{D}_{R_0} T^{\lambda}_{R_0}\|_p \stackrel{d}{=} \|\mathcal{D}^{(1,0)}_{[0,1]} \mathcal{B}\|_p$ when $F = G$ for $p \in [1,\infty]$. It therefore suffices to show $\|\mathcal{D}^{(1,0)}_{[0,1]} \mathcal{B}\|_p \ge_{\mathrm{S}} \|d\mathcal{D}_R T^{\lambda}_R\|_p$ for $p \in [1,\infty]$ and for any other $R \in \Theta_0$. If $R \in \Theta_0$ is strictly star-shaped, then from Theorem 2.1, $\|d\mathcal{D}_R T^{\lambda}_R\|_p = 0$ for $p \in [1,\infty]$ and hence $\|\mathcal{D}^{(1,0)}_{[0,1]} \mathcal{B}\|_p \ge_{\mathrm{S}} \|d\mathcal{D}_R T^{\lambda}_R\|_p$. If $R \in \Theta_0$ is non-strictly star-shaped, then for $p \in [1,\infty)$,

$$\begin{aligned} \|\mathcal{D}^{(1,0)}_{[0,1]} \mathcal{B}\|_p = \left[\int_0^1 \{\mathcal{D}^{(1,0)}_{[0,1]} \mathcal{B}(u)\}^p \, du\right]^{1/p} \ &\ge \ \left[\sum_k \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[0,1]} \mathcal{B}(u)\}^p \, du\right]^{1/p} \\ &\ge \ \left[\sum_k \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]} \mathcal{B}(u)\}^p \, du\right]^{1/p}. \end{aligned} \quad (2.1)$$

The first and second inequalities above hold because $\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}(u) \geq 0$ and $\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}(u) \geq \mathcal{D}^{(1,0)}_{[a_k,b_k]}\mathcal{B}(u) \geq 0$, for all $u \in [0,1]$. Because $\lambda \in (0,1)$ and $R'(a_k) \leq 1$ for all k, $\lambda R'(a_k) + (1-\lambda)R'(a_k)^2 \leq 1$ and the rightmost side of (2.1) is greater than or equal to

$$\left[\sum_k \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{p/2} \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]}\mathcal{B}(u)\}^p du\right]^{1/p} \stackrel{d}{=} \|d\mathcal{D}_R T_R^\lambda\|_p.$$

Therefore, for $R \in \Theta_0$ non-strictly star-shaped, we have $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p \geq_{\mathrm{S}} \|d\mathcal{D}_R T_R^\lambda\|_p$ for $p \in [1,\infty)$. Showing $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_\infty \geq_{\mathrm{S}} \|d\mathcal{D}_R T_R^\lambda\|_\infty$ for $R \in \Theta_0$ non-strictly star-shaped is analogous. $\square$

Proof of Theorem 2.3. When $R \in \Theta_1$, we redefine the functional operator $d\mathcal{D}_R : l([0,1]) \mapsto l([0,1])$ in Theorem 2.1 by

$$d\mathcal{D}_R h(u) = \begin{cases} -h(1), & \text{if } u = 1,\ R(u) < 1 \\ \max\{h(1), 0\} - h(1), & \text{if } u = 1,\ R(u) = 1 \\ \mathcal{L}_{S_k} h(u) - h(u), & \text{if } \exists k \text{ such that } u \in S_k \setminus \{1\} \\ 0, & \text{otherwise.} \end{cases}$$

The proof proceeds in the same manner as in Theorem 2.1. If $\mathcal{D}$ is not Hadamard directionally differentiable, one can use Skorohod's representation theorem and part (b) of Lemma A.7 in Appendix to obtain the result. $\square$

Proof of Theorem 2.4. For convenience, all limits stated in this proof assume that $\max\{m,n\} = O(r)$ and $n/(m+n) \to \lambda \in (0,1)$, as $r \to \infty$. We have independent random samples $X_1^{(r)}, X_2^{(r)}, ..., X_m^{(r)}$ and $Y_1^{(r)}, Y_2^{(r)}, ..., Y_n^{(r)}$ from $F^{(r)}$ and $G^{(r)}$, respectively. The sample ODC is $R_{mn}^{(r)} = F_m^{(r)}(G_n^{(r)})^{-1}$, where $F_m^{(r)}$ and $(G_n^{(r)})^{-1}$ are the empirical distribution and empirical quantile functions, respectively. Our test statistic is $M_{mn}^p = c_{mn}\|\mathcal{D}R_{mn}^{(r)}\|_p$. By the triangle inequality,

$$\mathrm{pr}_{R^{(r)} \in \Theta_1}(M_{mn}^p \geq c_{\alpha,p}) \geq \mathrm{pr}_{R^{(r)} \in \Theta_1}(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p < c_{mn}\|\mathcal{D}R^{(r)}\|_p - c_{\alpha,p})$$

for all $p \in [1, \infty]$. Therefore, to prove part (a), it suffices to show that $c_{mn} \|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p = O_P(1)$.

From Lemma A.3 in Appendix A, it follows that $\|\mathcal{M}R_{mn}^{(r)} - \mathcal{M}R^{(r)}\|_\infty \leq \|R_{mn}^{(r)} - R^{(r)}\|_\infty$, which implies $\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_\infty \leq 2\|R_{mn}^{(r)} - R^{(r)}\|_\infty$. Because L^p norms are dominated by the sup-norm, it therefore suffices to show $c_{mn}\|R_{mn}^{(r)} - R^{(r)}\|_\infty$ is $O_P(1)$. To accomplish this, we decompose $c_{mn}(R_{mn}^{(r)} - R^{(r)})$ into two parts:

$$c_{mn}(R_{mn}^{(r)} - R^{(r)}) = c_{mn}\{F_m^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G_n^{(r)})^{-1}\} \\ + c_{mn}\{F^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G^{(r)})^{-1}\}. \quad (2.2)$$

Define the two independent empirical processes

$$U_m(u) = \frac{1}{m}\sum_{i=1}^{m} I\{F^{(r)}(X_i^{(r)}) \leq u\}$$

and $V_n(u) = n^{-1}\sum_{i=1}^{n} I\{G^{(r)}(Y_i^{(r)}) \leq u\}$, for $0 \leq u \leq 1$. This allows us to rewrite $F_m^{(r)}$ as $U_m F^{(r)}$ and $F^{(r)}(G_n^{(r)})^{-1}$ as $R^{(r)}V_n^{(r)}$. Consequently, the two terms on the right-hand side of Equation (2.2) can be written as

$$c_{mn}\{F_m^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G_n^{(r)})^{-1}\} = c_{mn}[U_m\{F^{(r)}(G_n^{(r)})^{-1}\} - U\{F^{(r)}(G_n^{(r)})^{-1}\}] \quad (2.3)$$

and

$$c_{mn}\{F^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G^{(r)})^{-1}\} = c_{mn}(R^{(r)}V_n - R^{(r)}V), \quad (2.4)$$

where $U(\cdot)$ and $V(\cdot)$ both represent the cumulative distribution function of a uniform distribution on $[0, 1]$. These expressions allow us to unify all random samples (from different distributions) to be uniformly distributed.

We are now ready to show that the sup-norms of the right-hand sides of Equations (2.3) and (2.4) are uniformly bounded in probability. We begin with the uniform processes. From Theorem 3 in Komlós et al. (1975), there exist versions of independent standard Brownian bridges $\mathcal{B}_1^{(m)}$ and $\mathcal{B}_2^{(n)}$ such that, almost surely,

$$\|\sqrt{m}(U_m - U) - \mathcal{B}_1^{(m)}\|_\infty = o(m^{-1/2}(\log m)^2) \quad (2.5)$$

$$\|\sqrt{n}(V_n - V) - \mathcal{B}_2^{(n)}\|_\infty = o(n^{-1/2}(\log n)^2). \quad (2.6)$$

Because $\lim c_{mn}/(\lambda^{1/2}\sqrt{m}) = 1$, it is sufficient to say $\|c_{mn}(U_m - U) - \lambda^{1/2}\mathcal{B}_1^{(m)}\|_\infty = o(m^{-1/2}(\log m)^2)$ from Equation (2.5). Consequently, the sup-norm of the right-hand side of Equation (2.3) is less than or equal to

$$\|c_{mn}[U_m\{F^{(r)}(G_n^{(r)})^{-1}\} - U\{F^{(r)}(G_n^{(r)})^{-1}\}] - \lambda^{1/2}\mathcal{B}_1^{(m)}\{F^{(r)}(G_n^{(r)})^{-1}\}\|_\infty \\ + \|\lambda^{1/2}\mathcal{B}_1^{(m)}\{F^{(r)}(G_n^{(r)})^{-1}\}\|_\infty$$

which is less than or equal to

$$\|c_{mn}(U_m - U) - \lambda^{1/2}\mathcal{B}_1^{(m)}\|_\infty + \|\lambda^{1/2}\mathcal{B}_1^{(m)}\|_\infty = o(m^{-1/2}(\log m)^2) + O_P(1).$$

The $O_P(1)$ term arises because $\mathcal{B}_1^{(m)}$ is bounded with probability 1. Likewise, the $o(m^{-1/2}(\log m)^2)$ term comes from Equation (2.5). Therefore, we have shown that the sup-norm of the right-hand side of Equation (2.3), that is, $\|c_{mn}\{F_m^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G_n^{(r)})^{-1}\}\|_\infty = O_P(1)$.

For the right-hand side of Equation (2.4), we use the mean value theorem to write

$$R^{(r)}V_n(u) - R^{(r)}V(u) = \dot{R}^{(r)}(\tau_u)\{V_n(u) - V(u)\},$$

where $\dot{R}^{(r)}$ denotes the derivative of $R^{(r)}$ and where τ_u is between $V_n(u)$ and $V(u)$. Therefore,

$$\sup_{u\in[0,1]} \left|\sqrt{n}\{R^{(r)}V_n(u) - R^{(r)}V(u)\} - \dot{R}^{(r)}(\tau_u)\mathcal{B}_2^{(n)}(u)\right| \\ = \sup_{u\in[0,1]} \left|\dot{R}^{(r)}(\tau_u)[\sqrt{n}\{V_n(u) - V(u)\} - \mathcal{B}_2^{(n)}(u)]\right|$$

which is less than or equal to

$$\|\dot{R}^{(r)}\|_\infty\|\sqrt{n}(V_n - V) - \mathcal{B}_2^{(n)}\|_\infty = O(1)o(n^{-1/2}(\log n)^2) = o(n^{-1/2}(\log n)^2).$$

The $O(1)$ term above comes from the assumption that the derivative of $R^{(r)}$ is uniformly bounded for all r over $[0, 1]$. Likewise, the $o(n^{-1/2}(\log n)^2)$ term comes from Equation (2.6). Therefore, because $\lim c_{mn}/\{(1-\lambda)^{1/2}\sqrt{n}\} = 1$ and because $\mathcal{B}_2^{(n)}$ is bounded with probability 1, we have shown the sup-norm of the right-hand side of Equation (2.4), that is, $c_{mn}\|R^{(r)}V_n^{(r)} - R^{(r)}V\|_\infty = O_P(1)$. Finally, from Equation (2.2), we have $c_{mn}\|R_{mn}^{(r)} - R^{(r)}\|_\infty = O_P(1) + O_P(1) = O_P(1)$, which establishes part (a).

To prove part (b), let $q_{\beta,p}^{(r)}$ denote the $1-\beta$ quantile of the finite-sample distribution of $c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p$; i.e., $q_{\beta,p}^{(r)}$ solves $\mathrm{pr}_{R^{(r)}\in\Theta_1}(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p \le q_{\beta,p}^{(r)}) = 1-\beta$. We have already shown $c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p = O_P(1)$, so $\sup_r q_{\beta,p}^{(r)} \equiv q_{\beta,p} < \infty$. Therefore,

$$\liminf \mathrm{pr}_{R^{(r)}\in\Theta_1}(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p < q_{\beta,p}) \ge 1-\beta.$$

Set $\eta_p(\beta) = q_{\beta,p} + c_{\alpha,p}$. Whenever $\liminf c_{mn}\|\mathcal{D}R\|_p \ge \eta_p(\beta)$, it follows from the triangle inequality that $\liminf \mathrm{pr}_{R^{(r)}\in\Theta_1}(M_{mn}^p \ge c_{\alpha,p})$ is greater than or equal to

$$\begin{aligned}
&\liminf \mathrm{pr}_{R^{(r)}\in\Theta_1}(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p < c_{mn}\|\mathcal{D}R^{(r)}\|_p - c_{\alpha,p})\\
&\qquad \ge \liminf \mathrm{pr}_{R^{(r)}\in\Theta_1}(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p < q_{\beta,p}) \ge 1-\beta.
\end{aligned}$$

This completes the proof of part (b). $\square$

Chapter 3

Empirical-likelihood-based testing for positive quadrant dependence

Summary: Positive quadrant dependence (PQD) has attracted interest in statistics due to its practical importance in insurance, finance, reliability analysis, and actuarial science. This chapter develops solutions for two problems: testing independence versus PQD and testing PQD versus non-PQD between two random variables. Our approach is based on the empirical likelihood. It first localizes the tests according to the position of data points and then integrates the local test statistics to form global test statistics. We derive asymptotic distributions under the corresponding null hypotheses and suggest limiting least favorable configurations to construct rejection regions. Simulation is conducted to demonstrate the finite sample performance of our methods and to support the theoretical results. Finally, we illustrate our methods using stock price data sets.

3.1 Introduction

In many applications, it is important to characterize the dependence structure between two random variables X and Y. Positive dependence exists when X and Y tend to simultaneously provide small values (large values as well) and is commonly described numerically by the correlation between these two random variables. A stronger form of this type of dependence, positive quadrant dependence (PQD), was proposed by Lehmann (1966). This dependence holds between X and Y if

$$\mathrm{pr}(X \leq x, Y \leq y) \geq \mathrm{pr}(X \leq x)\mathrm{pr}(Y \leq y), \tag{3.1}$$

for all $x, y \in \mathbb{R}$. In other words, the probability of X and Y being simultaneously small is at least as large as it would be when X and Y are independent.

In this dissertation, we are interested in PQD because it plays a key role in a variety of applications. For example, risk management experts are often interested in comparing the prices of an insurance premium containing two assets with values X and Y under PQD to the same premium under independence. Having this knowledge can help these experts establish more suitable selling prices with the assets combined (Dhaene and Goovaerts, 1996; Denuit and Scaille, 2004). In reliability theory, consider the survival times X and Y of two components arranged in a parallel system. If the failure of one component leads to an earlier failure of the other, then X and Y satisfy PQD. In this case, incorrectly assuming independence would overestimate the reliability of a system (Lai, 2003; Lai and Xie, 2006).

The practical importance of PQD motivates the need to develop statistical procedures to test for it. Denote by H the joint distribution function of (X, Y) with marginals F and G, respectively. The definition of PQD in (3.1) is equivalent to $H(x, y) \geq F(x)G(y)$ for all $x, y \in \mathbb{R}$. Thus, in this chapter, we consider the following three hypotheses

$$\begin{aligned}
\mathcal{H}_0 &: H(x, y) = F(x)G(y), \quad \text{for all } x, y \\
\mathcal{H}_1 &: H(x, y) \geq F(x)G(y), \quad \text{for all } x, y \\
\mathcal{H}_2 &: H(x, y) \neq F(x)G(y), \quad \text{for some } x, y.
\end{aligned}$$

We first develop tests for PQD; i.e., $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$. Then we propose new goodness-of-fit tests for PQD; i.e., $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$.

Existing methods for the two tests; i.e., $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$ and $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$, can be best characterized as falling into two groups: those that are parametrization-based and those that are distance-based. The spaces consisting of all possible H in $\mathcal{H}_0$, $\mathcal{H}_1 - \mathcal{H}_0$, or $\mathcal{H}_2 - \mathcal{H}_1$ are each infinite dimensional. The first collection of methods parametrizes these infinite dimensional spaces into those that are finite-dimensional and then tests the transformed hypotheses. Distance-based methods focus on the departure between estimates of H under the null or the alternative hypotheses. Because $\mathcal{H}_1$ is also equivalent to $C(u, v) \geq \Pi(u, v)$ where C is a copula such that $C\{F(x), G(y)\} = H(x, y)$ for all $x, y \in \mathbb{R}$, and $\Pi(u, v) = uv$ is the independence copula (Sklar, 1959), many distance-based methods actually measure the distance between an estimate of $C(u, v)$ and $\Pi(u, v)$ over $u, v \in [0, 1]$.

When testing $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$, parametrization-based methods include Kochar and Gupta (1987) and Janic-Wròbelwska et al. (2004). Kochar and Gupta (1987) tested a higher-order Kendall's tau coefficient to be equal to or greater than zero with a given order. Their coefficients include the commonly used Kendall's tau as a special case when the order is one. In this case, they used a one-dimensional parameter to characterize the infinite-dimensional hypotheses. By doing so, it actually enlarges $\mathcal{H}_1$ and consequently reduces the power of the test. Janic-Wròbelwska et al. (2004) restricted attention to a finite-dimensional subset of $\mathcal{H}_1 - \mathcal{H}_0$ which is determined by a parametrization using a normalized Legendre polynomial approximation. However, if this approximation is inadequate for the true configuration of X and Y that satisfies PQD, their methods may lose power. One representative example of a distance-based method for $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$ comes from Denuit and Scaille (2004), where estimates of the departure of H or the copula from independence were used as test statistics. Their methods require consistent estimates of a covariance matrix, partial derivatives of H, and densities of F and G. Not surprisingly, any bias in these estimates can affect the power of these tests.

When testing $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$, Ledwina and Wyłupek (2014) proposed a test in the same spirit of Janic-Wròbelwska et al. (2004). However, instead of using a polynomial approximation, they used a set of non-decreasing step functions constructed on equally-spaced grid points between 0 and 1 to approximate the null hypothesis $\mathcal{H}_1$. In other words, their testing procedure concentrated on a finite-dimensional subset of the null hypothesis. When the true PQD relationship is not in the subset, the Type I error probability will exceed the significance level. Regarding this test, of more interest to us are the distance-based methods proposed by Scaillet (2005), Gijbels et al. (2010), and Gijbels and Sznajder (2013). Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling type test statistics were constructed according to different functional distances of an estimated copula from $\mathcal{H}_1$ (either empirically or by using nonparametric kernel smoothing) and critical values were determined using bootstrapping. However, copula estimation can be subjective depending on the selection of the kernel function and the bandwidths. Further, their proposed resampling techniques require sampling from a restricted estimator of the copula C under PQD. Such a process can be time-consuming.

In this dissertation, we use empirical likelihood (EL) to construct test statistics for the two PQD problems. The EL method was originally proposed by Owen (1990) and was quickly recognized as an excellent method for constructing confidence regions in many settings. The numerous advantages of EL have been described, which include enjoying nonparametric flexibility, maintaining the likelihood-ratio-based efficiency, automatically respecting the range of the parameter space, and its Bartlett-correctability. The use of EL to test nonparametric hypotheses was developed by Einmahl and McKeague (2003), where they considered testing for symmetry, a change point, independence, and exponentiality. Following in a similar spirit, El Barmi and McKeague (2013) extended the utility of EL to test for stochastic ordering. These works demonstrate the advantages of EL-based test statistics over traditional ones such as the Cramér-von Mises statistics. Moreover, when compared to the aforementioned two groups of approaches, EL-based methods do not reparametrize the hypotheses into finite dimensions, and EL does not require one to specify which functional distance should be used. These advantages inspire us to derive new tests for PQD.

To test $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$ and $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$, we first localize the test and derive EL-based local test statistics. The final global test statistics are obtained by integrating the local test over all data points. Asymptotic properties of our test statistics under corresponding null hypotheses are derived. See Sections 3.2 and 3.3 for details. For testing $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$, we show our test statistics are distribution-free which enable us to find critical values. When compared to the two-sided test for independence in Einmahl and McKeague (2003), our test is more powerful if the dependence is truly PQD. For testing $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$, we show the independence structure provides the least favorable configuration which is used to determine rejection regions. Simulation shows our EL-based tests outperform the distance-based procedures provided by Scaillet (2005), Gijbels et al. (2010), and Gijbels and Sznajder (2013). All numerical results are presented in Section 3.4. Finally, we apply our new tests to a stock price data set in Section 3.5 and conclude this chapter with a discussion in Section 3.6.

3.2 Testing $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$

Suppose that $\{(X_i, Y_i)\}_{i=1}^n$ is a random sample from a continuous joint distribution H with marginal distributions F and G. The goal of this section is to test $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$. Following the method outlined by Einmahl and McKeague (2003), our procedure first localizes the hypotheses at a given point to derive a local EL-based test statistic. Pick any point $(x, y)' \in \mathbb{R}^2$. The local versions of $\mathcal{H}_0$ and $\mathcal{H}_1$ at $(x, y)'$ are denoted by $\mathcal{H}_0^{(x,y)} : H(x, y) = F(x)G(y)$ and $\mathcal{H}_1^{(x,y)} : H(x, y) \geq F(x)G(y)$, respectively. One could interpret $\mathcal{H}_0^{(x,y)}$ to be the independence between X and Y at $(x, y)'$ and $\mathcal{H}_1^{(x,y)}$ to be the local PQD between X and Y at $(x, y)'$. To test $\mathcal{H}_0^{(x,y)}$

versus $\mathcal{H}_1^{(x,y)} - \mathcal{H}_0^{(x,y)}$, we use the localized empirical likelihood ratio

$$R_{01}(x,y) = \frac{\sup\{L(\tilde{H}) : \tilde{H}(x,y) = \tilde{F}(x)\tilde{G}(y)\}}{\sup\{L(\tilde{H}) : \tilde{H}(x,y) \geq \tilde{F}(x)\tilde{G}(y)\}},$$

where $L(\tilde{H})$ is the empirical likelihood of observing the samples from a possible joint distribution $\tilde{H}$ of (X,Y); i.e., the product of the probability masses that were assigned at data points by $\tilde{H}$. Mathematically, $L(\tilde{H})$ could also be interpreted as $\prod_{i=1}^n \{\tilde{H}(X_i,Y_i) - \tilde{H}(X_i-,Y_i-)\}$, where $\tilde{H}(a-,b-) = \lim_{\delta_1\downarrow 0,\delta_2\downarrow 0} \tilde{H}(a-\delta_1, b-\delta_2)$. For notational convention, we let $\sup\emptyset = 0$ and $0/0 = 1$. Note that the numerator of $R_{01}(x,y)$ is the supremum of empirical likelihoods that satisfy the local null hypothesis $\mathcal{H}_0^{(x,y)}$ and the denominator is the supremum of those satisfying $\mathcal{H}_1^{(x,y)}$.

To explore $R_{01}(x,y)$, define four closely related regions based on $(x,y)'$: $A_{11} = (-\infty, x] \times (-\infty, y]$, $A_{12} = (-\infty, x] \times (y, \infty)$, $A_{21} = (x,\infty) \times (-\infty, y]$, and $A_{22} = (x,\infty) \times (y,\infty)$. Let P_n be the empirical measure corresponding to the observed data. For example, $P_n(A_{11}) = n^{-1}\sum_{i=1}^n I\{(X_i,Y_i) \in A_{11}\}$ where $I\{\cdot\}$ is the indicator function; i.e., $P_n(A_{11})$ is the proportion of samples falling in region A_{11}. We further denote by H_n the empirical estimator of H, by F_n and G_n the empirical estimators of F and G, respectively. Using the Lagrange multiplier method, we are able to obtain an explicit formula of the empirical likelihood ratio $R_{01}(x,y)$; i.e., if $H_n(x,y) \leq F_n(x)G_n(y)$, then $R_{01}(x,y) = 1$; otherwise,

$$\begin{aligned} R_{01}(x,y) = & \left(\frac{F_n(x)G_n(y)}{P_n(A_{11})}\right)^{nP_n(A_{11})} \left(\frac{F_n(x)\{1-G_n(y)\}}{P_n(A_{12})}\right)^{nP_n(A_{12})} \\ & \times \left(\frac{\{1-F_n(x)\}G_n(y)}{P_n(A_{21})}\right)^{nP_n(A_{21})} \left(\frac{\{1-F_n(x)\}\{1-G_n(y)\}}{P_n(A_{22})}\right)^{nP_n(A_{22})}. \end{aligned}$$

Detailed derivations are presented in Appendix B.

Consequently, a local test statistic can be obtained by $-2\ln R_{01}(x,y)$, where we set $0 \times \ln 0 = 0$ for notational convenience. This test statistic is nonnegative and takes positive values only when $H_n(x,y) > F_n(x)G_n(y)$. One can view $-2\ln R_{01}(x,y)$ as a measure of a pointwise departure of H from the local independence at $(x,y)'$. Under

$\mathcal{H}_0^{(x,y)}$, a second-order Taylor expansion, combined with the central limit theorem and the continuous mapping theorem, yields that $-2\ln R_{01}(x,y)$ converges in distribution to $Z^2 I(Z>0)$ when $n\to\infty$, where Z follows the standard normal distribution. Note that the limiting distribution does not depend on the location $(x,y)'$ nor the distribution functions H, F, or G. This asymptotic property reveals another advantage of the EL-based method; i.e., test statistics are automatically self-standardized.

Once we obtain a test statistic at $(x,y)'$, the next step is to integrate it over $\mathbb{R}^2$ to obtain global test statistics. Towards this end, we consider two ways of integration and denote them by EL_1 and EL_2, that is,

$$\begin{aligned}
\mathrm{EL}_1 &= \int_{(x,y)\in\mathbb{R}^2} -2\ln R_{01}(x,y)\,dF_n(x)dG_n(y) = -\frac{2}{n^2}\sum_{i=1}^n\sum_{j=1}^n \ln R_{01}(X_i,Y_j)\\
\mathrm{EL}_2 &= \int_{(x,y)\in\mathbb{R}^2} -2\ln R_{01}(x,y)\,dH_n(x,y) = -\frac{2}{n}\sum_{i=1}^n \ln R_{01}(X_i,Y_i).
\end{aligned}$$

Our first test statistic, EL_1, integrates the local test statistics with respective to F_nG_n which is a version of H_n when $\mathcal{H}_0$ is true. An equal weight $1/n^2$ is put on the grid points (X_i,Y_j) for all $1\le i,j\le n$. This integration was suggested by Einmahl and McKeague (2003). The second test statistic EL_2 integrates the local test statistic with respective to H_n. An equal weight $1/n$ is put on the data points (X_i,Y_i) for $1\le i\le n$. For each test statistic, large values are evidence against $\mathcal{H}_0$. To determine critical values, we studied the asymptotic behavior of EL_1 and EL_2 under the null hypothesis $\mathcal{H}_0$. The results are stated in the following theorem.

Theorem 3.1. *Under $\mathcal{H}_0$; i.e., when X and Y are independent, we have*

$$\mathrm{EL}_1 \xrightarrow{d} \int_0^1\int_0^1 \frac{\{[\mathcal{B}_s(u,v)-u\mathcal{B}_s(1,v)-v\mathcal{B}_s(u,1)]_+\}^2}{u(1-u)v(1-v)}\,du\,dv,$$

as $n\to\infty$, where $\xrightarrow{d}$ means convergence in distribution, $[a]_+=\max\{0,a\}$, and $\mathcal{B}_s$ is a mean-zero Gaussian process on $[0,1]^2$ with covariance structure

$$cov\{\mathcal{B}_s(u_1,v_1),\mathcal{B}_s(u_2,v_2)\} = \min\{u_1,u_2\}\min\{v_1,v_2\} - u_1v_1u_2v_2,$$

for all (u_1, v_1) *and* (u_2, v_2) *in* $[0,1]^2$. *This result also holds for* EL_2*; i.e.,* EL_1 *and* EL_2 *are asymptotically equivalent in distribution.*

A proof of this theorem can be obtained by making a straightforward modification to the proof in Einmahl and McKeague (2003). For finding critical values, an immediate thought is to approximate the limiting distribution by simulating the Gaussian process $\mathcal{B}_s$. However, a second inspection of Theorem 3.1 tells us that the limiting distribution is distribution free; i.e., it does not depend on H. The following proposition further proves that both test statistics are distribution free in finite sample cases.

Proposition 3.2. *When n is finite, both* EL_1 *and* EL_2 *are marginal distribution free; i.e., their distributions do not depend on F or G. Further, when $\mathcal{H}_0$ is true, they are both distribution free.*

With this proposition, it suffices to consider one specific joint distribution under $\mathcal{H}_0$ to determine critical values when n is finite. For simplicity, we choose the uniform distribution with support $(0,1)^2$, denoted by uniform$(0,1)^2$. We generated $100,000$ Monte Carlo samples of size n from uniform$(0,1)^2$. Based on each sample, we computed the values of both test statistics. For each test statistic, the estimated critical value at significance level α was chosen to be the $1-\alpha$ quantile of the $100,000$ values of the test statistic. We present our estimates for selected significance levels in Table B.1. We reject the null hypothesis $\mathcal{H}_0$ at level α when the value of the test statistic exceeds the corresponding critical value.

3.3 Testing $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$

Testing $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$ is more complicated than testing $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$ due to the composite structure of the null hypothesis $\mathcal{H}_1$. However, the construction of the EL-based test statistics still follows the same steps. The first step is to test a localized version of the hypotheses. Fix $(x,y)' \in \mathbb{R}^2$. The local hypotheses at $(x,y)'$ are $\mathcal{H}_1^{(x,y)} : H(x,y) \geq F(x)G(y)$ versus $\mathcal{H}_2^{(x,y)} - \mathcal{H}_1^{(x,y)} : H(x,y) < F(x)G(y)$. Similarly as in Section 3.2, we define the local empirical likelihood ratio $R_{12}(x,y)$ by

$$R_{12}(x,y) = \frac{\sup\{L(\tilde{H}) : \tilde{H}(x,y) \geq \tilde{F}(x)\tilde{G}(y)\}}{\sup\{L(\tilde{H})\}},$$

where the numerator is the supremum of empirical likelihoods that satisfying the local PQD at $(x,y)'$, and the denominator is the supremum of those under no restriction. In Appendix B, we derive that $R_{12}(x,y) = 1$ if $H_n(x,y) \geq F_n(x)G_n(y)$; otherwise,

$$\begin{aligned} R_{12}(x,y) = & \left(\frac{F_n(x)G_n(y)}{P_n(A_{11})}\right)^{nP_n(A_{11})} \left(\frac{F_n(x)\{1-G_n(y)\}}{P_n(A_{12})}\right)^{nP_n(A_{12})} \\ & \times \left(\frac{\{1-F_n(x)\}G_n(y)}{P_n(A_{21})}\right)^{nP_n(A_{21})} \left(\frac{\{1-F_n(x)\}\{1-G_n(y)\}}{P_n(A_{22})}\right)^{nP_n(A_{22})}. \end{aligned}$$

One can see the similarity between $R_{12}(x,y)$ and $R_{01}(x,y)$. When H_n violates the corresponding local null hypothesis, they both take the same ratio; otherwise, they take the value 1.

Again, we choose $-2\ln R_{12}(x,y)$ as the local test statistic. It is nonnegative and takes positive values when H_n violates the local PQD assumption at $(x,y)'$; i.e., $H_n(x,y) < F_n(x)G_n(y)$. We view it as a measure of a pointwise departure of H_n from $\mathcal{H}_1^{(x,y)}$. Under $\mathcal{H}_1^{(x,y)}$, the limiting distribution of $-2\ln R_{12}(x,y)$ depends on whether H satisfies the local PQD assumption strictly or not; i.e., whether $H(x,y) = F(x)G(y)$ or $H(x,y) > F(x)G(y)$. If $H(x,y) = F(x)G(y)$, the limiting distribution of $-2\ln R_{12}(x,y)$ is the same as the one of $-2\ln R_{01}(x,y)$ under $\mathcal{H}_0^{(x,y)}$; i.e., it converges in distribution to $Z^2 I(Z<0)$ as $n \to \infty$. On the other hand, if $H(x,y) < F(x)G(y)$, we find that $-2\ln R_{12}(x,y)$ converges to zero in probability. In summary, we conclude that, as $n \to \infty$,

$$-2\ln R_{12}(x,y) \xrightarrow{d} \begin{cases} Z^2 I(Z<0), & \text{if } H(x,y) = F(x)G(y) \\ 0, & \text{if } H(x,y) > F(x)G(y). \end{cases}$$

Finally, we aggregate the local test statistics to form global test statistics. Similarly as in the construction of EL_1 and EL_2, we define

$$\mathrm{EL}_3 = \int_{(x,y)\in\mathbb{R}^2} -2\ln R_{12}(x,y)\, dF_n(x)dG_n(y) = -\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\ln R_{12}(X_i,Y_j)$$

$$\mathrm{EL}_4 = \int_{(x,y)\in\mathbb{R}^2} -2\ln R_{12}(x,y)\, dH_n(x,y) = -\frac{2}{n}\sum_{i=1}^{n}\ln R_{12}(X_i,Y_i)$$

to be our final test statistics. Large values of EL_3 and EL_4 are evidence against $\mathcal{H}_1$. However, finding suitable critical values is not as straightforward as in Section 3.2. This is because under the null hypothesis $\mathcal{H}_1$, neither of the two test statistics are distribution free which makes it difficult to control Type I error probability. Fortunately, we noticed that the limiting distributions of local test statistics have a nice form. In this light, we proceed to study the asymptotic behavior of the global test statistics.

From the limiting distribution of the local test statistics, it is expected that EL_3 and EL_4 depend on the regions that H satisfy the local PQD assumption strictly or not. However, the proofs here are more complicated because the asymptotic behavior of local test statistics are very different (converges in distribution to zero or to $Z^2 I(Z<0)$) so that standard empirical process theory used in Einmahl and McKeague (2003) cannot be directly applied here. So we state the limiting distribution of EL_3 and EL_4 as a conjecture below. Further, similar to Theorem 3.1, the limiting distribution should depend on the same Gaussian process $\mathcal{B}_s$ defined in Theroem 3.1 as well.

Conjecture 3.1. Under $\mathcal{H}_1$, as $n\to\infty$, both EL_3 and EL_4 converge in distribution to

$$\int_{\{(u,v):C(u,v)=\Pi(u,v)\}} \frac{\{[\mathcal{B}_s(u,v)-u\mathcal{B}_s(1,v)-v\mathcal{B}_s(u,1)]_-\}^2}{uv(1-u)(1-v)}\, du\, dv.$$

where $\mathcal{B}_s$ is defined in Theorem 3.1 and $[a]_- = \min\{0,a\}$.

It turns out that from Conjecture 3.1, one can see that the limiting distribution does not depend on the marginal distributions F or G. It solely depends on the joint distribution function H through the region $\{(u,v) : C(u,v) = \Pi(u,v)\}$ where the integration is carried out. Further, the integrand does not depend on H and is always nonnegative. Thus, the limiting distribution reaches its stochastic maximum when $C(u,v) = \Pi(u,v)$. We state this observation as a corollary below.

Corollary 3.3. *Under $\mathcal{H}_1$, for any constant c, we have*

$$\lim_{n\to\infty} \mathrm{pr}(\mathrm{EL}_k \geq c) \leq \mathrm{pr}\left(\int_0^1 \int_0^1 \frac{\{[\mathcal{B}_s(u,v) - u\mathcal{B}_s(1,v) - v\mathcal{B}_s(u,1)]_-\}^2}{uv(1-u)(1-v)}\, du\, dv > c\right),$$

for $k = 3, 4$.

Corollary 3.3 demonstrates that independence between X and Y; i.e., when $C = \Pi$, yields the least favorable configuration for testing $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$. In other words, controlling the Type I error probability by assuming X and Y are independent uniform$(0,1)$ random variables is sufficient to control the Type I error probability under any other configuration of H that satisfies $\mathcal{H}_1$. Thus, it allows us to take the same approach as in Section 3.2 to estimate critical values. We generated $100,000$ Monte Carlo samples of size n from uniform$(0,1)^2$ and calculated the test statistics using each sample. At significance level α, the $1-\alpha$ quantile of the $100,000$ values was taken to estimate the critical values. We provide selected results in Table B.2 in Appendix B. The null hypothesis $\mathcal{H}_1$ should be rejected at level α when test statistics are larger than the corresponding critical values.

3.4 Simulation results

In this section, we illustrate the finite sample performance of the proposed EL-based testing procedures.

3.4.1 $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$

We first consider testing $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$ and compare our EL-based tests with a procedure proposed by Einmahl and McKeague (2003) and three traditional distance-based methods. Einmahl and McKeague (2003) proposed a test for $\mathcal{H}_0$ versus $\mathcal{H}_2 - \mathcal{H}_0$; we denote their test statistic by

$$\mathrm{EM} = -\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \ln R_{02}(X_i, Y_j),$$

where $R_{02}(x,y) = R_{01}(x,y) \times R_{12}(x,y)$. They suggested rejecting $\mathcal{H}_0$ when EM is large. Comparing to this test enables us to see whether our tests improve the power when X and Y when are actually PQD. To reveal the advantages of the EL-based method over traditional distance-based testing procedures, we considered three commonly used distances: Kolmogorov-Smirnov (KS), Cramér-von Mises (CvM), and Anderson-Darling (AD) distances. Recall that $\mathcal{H}_1$ is equivalent to $C(u,v) \geq \Pi(u,v)$ and $\mathcal{H}_0$ is equivalent to $C(u,v) = \Pi(u,v)$ for all $u, v \in [0,1]$. Intuitively, a positive departure of C from Π, measured by $[C(u,v) - uv]_+$, gives evidence of rejecting $\mathcal{H}_0$. Using the empirical copula $\widehat{C}_n$ proposed by Deheuvels (1979), where

$$\widehat{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{n}{n+1}F_n(X_i) \leq u, \frac{n}{n+1}G_n(Y_i) \leq v\right),$$

we construct KS, CvM, AD-distance-based test statistics by

$$\begin{aligned}
\mathrm{KS}_1 &= \sqrt{n} \sup_{(u,v)\in[0,1]^2} [\widehat{C}_n(u,v) - uv]_+ \\
\mathrm{CvM}_1 &= n\int_{[0,1]^2} [\widehat{C}_n(u,v) - uv]_+^2 \, d\widehat{C}_n(u,v) \\
\mathrm{AD}_1 &= n\int_{[0,1]^2} \frac{[\widehat{C}_n(u,v) - uv]_+^2}{uv(1-u)(1-v)} \, d\widehat{C}_n(u,v),
\end{aligned}$$

respectively. Clearly, large values of KS_1, CvM_1, and AD_1 are evidence to reject $\mathcal{H}_0$. Note that all the four test statistics are distribution free when $\mathcal{H}_0$ holds. Thus, we estimate critical values for each of them via the same method introduced in Section 3.2; i.e., generating $100,000$ Monte Carlo samples of size n from uniform$(0,1)^2$ and choosing the $1-\alpha$ quantile of the $100,000$ values of the test statistic as the critical value for a significance level α. We provide estimated critical values in Table B.1.

For the purpose of comparison, we consider random samples $\{(X_i, Y_i)\}_{i=1}^n$ from a bivariate normal distribution where the marginals are standard normal and the correlation coefficient is ρ. In this case, X and Y are positive quadrant dependent if and only if $\rho > 0$, and X and Y are independent if and only if $\rho = 0$. We use $\rho = 0$ to compare sizes and $\rho \in \{0.25, 0.50\}$ to compare powers of all considered testing procedures. The sample size n is chosen to be in $\{10, 30, 50, 100\}$. For each (n, ρ), we generated a sample of size n from the bivariate normal distribution with correlation ρ and conducted all six testing procedures. Rejection rates were estimated from $1,000$ replications and are presented in Table 3.1.

Table 3.1: Estimated probability of rejecting $\mathcal{H}_0$ with different sample size configurations and $\alpha = 0.05$ for testing procedures EL_1, EL_2, EM, KS_1, CvM_1, and AD_1 when samples of size n are generated from a bivariate normal distribution with the marginal being standard normal and correlation coefficient being ρ, where n is considered to be in $\{10, 30, 50, 100\}$ and ρ takes values of 0, 0.25, and 0.50. Note that, when $\rho = 0$, $\mathcal{H}_0$ is true; when $\rho = 0.25$ and 0.50, $\mathcal{H}_1 - \mathcal{H}_0$ is true.

$\rho = 0$	EL_1	EL_2	EM	KS_1	CvM_1	AD_1
$n = 10$	0.051	0.051	0.061	0.057	0.048	0.051
$n = 30$	0.045	0.053	0.056	0.046	0.053	0.041
$n = 50$	0.050	0.055	0.056	0.050	0.049	0.053
$n = 100$	0.046	0.052	0.047	0.055	0.058	0.052
$\rho = 0.25$	EL_1	EL_2	EM	KS_1	CvM_1	AD_1
$n = 10$	0.172	0.158	0.109	0.164	0.157	0.153
$n = 30$	0.384	0.361	0.246	0.267	0.360	0.294
$n = 50$	0.528	0.509	0.356	0.354	0.468	0.423
$n = 100$	0.771	0.769	0.640	0.607	0.744	0.688
$\rho = 0.50$	EL_1	EL_2	EM	KS_1	CvM_1	AD_1
$n = 10$	0.394	0.366	0.249	0.332	0.377	0.337
$n = 30$	0.866	0.847	0.748	0.681	0.840	0.777
$n = 50$	0.968	0.965	0.930	0.872	0.958	0.941
$n = 100$	1.000	1.000	1.000	0.995	1.000	1.000

When $\rho = 0$, we can see from Table 3.1 that the sizes of all testing procedures attain the nominal level 0.05 within the 99% confidence margin of error (which is

0.018). This demonstrates that the method we used to estimate critical values works well. When $\rho > 0$, the power of each method increases as expected as n increases. These patterns give evidence of consistency for each method. Comparing our EL-based tests with the distance-based ones, we can see that CvM_1 is the most powerful distance-based testing procedure, but both of EL_1 and EL_2 have larger powers than CvM_1. Such findings indicate the potential advantages of our EL-based tests over traditional distance-based methods. When comparing to EM, because PQD truly holds at $\rho = 0.25$ and 0.50, we see that both EL_1 and EL_2 improve the power of testing for independence as expected. Finally, we see little differences between the powers of EL_1 and EL_2 when n is large. But for small sample sizes, EL_1 does confer slightly larger power.

3.4.2 $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$

We now focus on the test of $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$ and compare our EL-based testing procedures with distance-based methods proposed by Gijbels et al. (2010) and Gijbels and Sznajder (2013). Before presenting the comparison, we briefly introduce their methods. Because the null hypothesis is now $\mathcal{H}_1$, the departure of H from $\mathcal{H}_1$ could be measured by $[\widehat{C}_n(u,v)-uv]_-$ instead of by $[\widehat{C}_n(u,v)-uv]_+$. Utilizing this measure, Gijbels et al. (2010) and Gijbels and Sznajder (2013) considered the same types of distances that were used in Section 3.4.1; i.e., KS, CvM, and AD distances. We denote the three test statistics subsequently by

$$\begin{aligned}
\mathrm{KS}_2 &= -\sqrt{n} \inf_{(u,v)\in[0,1]^2} [\widehat{C}_n(u,v) - uv]_- \\
\mathrm{CvM}_2 &= n \int_{[0,1]^2} [\widehat{C}_n(u,v) - uv]_-^2 \, d\widehat{C}_n(u,v) \\
\mathrm{AD}_2 &= n \int_{[0,1]^2} \frac{[\widehat{C}_n(u,v) - uv]_-^2}{uv(1-u)(1-v)} \, d\widehat{C}_n(u,v).
\end{aligned}$$

Large values of these test statistics provide evidence against $\mathcal{H}_1$. Similar to the difficulty met in Section 3.3, distributions of their test statistics under $\mathcal{H}_1$ depend on H

so that Type I error probabilities are less tractable. Gijbels et al. (2010) suggested estimating critical values by assuming $C = \Pi$. Following their suggestion, we generated $100,000$ Monte Carlo samples of size n from uniform$(0,1)^2$ which were then used to estimate critical values. The results are provided in Table B.2 in Appendix. Rather than treating Π as the least favorable configuration, Gijbels and Sznajder (2013) proposed a restricted bootstrapping method to estimate data-driven critical values and corresponding p-values so that larger powers could be achieved. To avoid notational confusion, we let KS_3, CvM_3, and AD_3 denote the bootstrapping versions of KS_2, CvM_2, and AD_2, respectively.

We first consider a bivariate normal distribution family with correlation ρ and marginals being standard normal, as in Section 3.4.1. Note that X and Y are not PQD if and only if $\rho < 0$; X and Y are PQD if and only if $\rho \geq 0$. Consequently, we choose $\rho \in \{0, 0.1, \ldots, 0.9\}$ to examine sizes and $\rho \in \{-0.9, -0.8, \ldots, -0.1\}$ to compare powers. To see the role of the sample size, we let $n \in \{10, 30, 50, 100\}$. For each (n, ρ), we generate $1,000$ samples of size n from the bivariate normal distribution with correlation ρ. Based on each sample, we apply the eight methods to conduct the test at significance level $\alpha = 0.05$, where for methods involving bootstrapping, $1,000$ bootstrap samples are used to estimate data-driven critical values. We summarize the estimated rejection rates for $\rho \in \{0, -0.25, -0.50\}$ in Table 3.2.

From Table 3.2, we see that when $\rho = 0$, the estimated sizes are all at the nominal level and within the 99% confidence margin of error of 0.018. This is expected for methods KS_2, CvM_2, and AD_2, because all select critical values using uniform$(0,1)^2$. The estimated sizes also suggest restricted bootstrapping works well for KS_3, CvM_3, and AD_3 when X and Y are independent. When PQD is violated; i.e., where $\rho \in \{-0.25, -0.50\}$, powers of all testing procedures quickly increase and approach one as the sample size increases. Among all distance-based methods, restricted bootstrapping can improve the power; e.g., AD_3 has better performance

Table 3.2: Estimated probability of rejecting $\mathcal{H}_1$ with different sample size configurations and $\alpha = 0.05$ for testing procedures EL_3, EL_4, KS_2, CvM_2, AD_2, KS_3, CvM_3, and AD_3 when samples of size n are generated from a bivariate normal distribution with the marginal being standard normal and correlation coefficient being ρ, where n is considered to be in $\{10, 30, 50, 100\}$ and ρ takes values of 0, -0.25, and -0.50. Note that, when $\rho = 0$, $\mathcal{H}_1$ is true; when $\rho = -0.25$ and -0.50, $\mathcal{H}_2 - \mathcal{H}_1$ is true.

$\rho = 0$	EL_3	EL_4	KS_2	CvM_2	AD_2	KS_3	CvM_3	AD_3
$n = 10$	0.062	0.063	0.075	0.065	0.064	0.052	0.067	0.068
$n = 30$	0.043	0.044	0.045	0.045	0.041	0.038	0.034	0.033
$n = 50$	0.049	0.045	0.043	0.048	0.043	0.043	0.045	0.045
$n = 100$	0.049	0.049	0.047	0.045	0.050	0.055	0.049	0.048
$\rho = -0.25$	EL_3	EL_4	KS_2	CvM_2	AD_2	KS_3	CvM_3	AD_3
$n = 10$	0.182	0.172	0.116	0.158	0.166	0.150	0.147	0.158
$n = 30$	0.322	0.307	0.238	0.268	0.272	0.256	0.291	0.301
$n = 50$	0.502	0.480	0.332	0.434	0.453	0.340	0.439	0.457
$n = 100$	0.784	0.765	0.601	0.733	0.764	0.595	0.713	0.749
$\rho = -0.50$	EL_3	EL_4	KS_2	CvM_2	AD_2	KS_3	CvM_3	AD_3
$n = 10$	0.409	0.342	0.269	0.335	0.343	0.334	0.327	0.342
$n = 30$	0.849	0.813	0.706	0.800	0.818	0.684	0.773	0.793
$n = 50$	0.970	0.958	0.844	0.940	0.950	0.868	0.943	0.951
$n = 100$	1.000	1.000	0.995	1.000	1.000	0.996	1.000	1.000

than AD_2. But overall, our proposed EL-based tests have the largest powers in every setting we considered.

We further visualize the comparison. Rejection rates of each testing procedure at $\rho \in \{-0.5, -0.4, \dots, 0.1, 0.2\}$ with sample size $n = 100$ at $\alpha = 0.05$ are connected in Figure 3.1. Results at $\rho \notin \{-0.5, -0.4, \dots, 0.1, 0.2\}$ are omitted because they are zero when $\rho \geq 0.3$ and one when $\rho \leq -0.6$. We can see that the curves of EL_3 and EL_4 are around 0.05 when $\rho = 0$ and decrease when ρ increases. This fact is in agreement with Corollary 3.3; i.e., independence between X and Y yields the least favorable configuration. Moreover, it seems that this least favorable configuration also holds for all the distance-based methods, even for the ones where bootstrapping was used. Finally, we would like to point it out again that our EL-based tests have dominating powers among all methods. Between the two EL-based tests, EL_3 is slightly better.

Beyond the bivariate normal distributions, we further consider two different de-

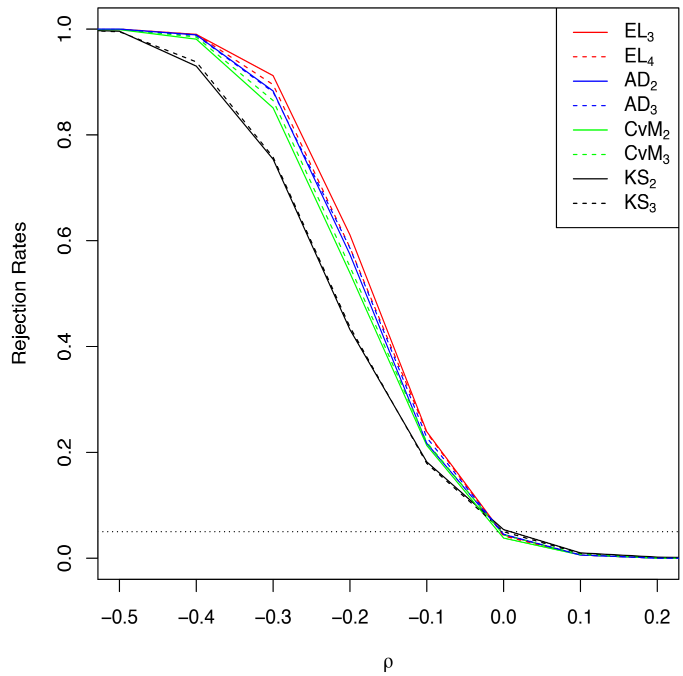

Figure 3.1: Rejection rates of testing procedures EL_3, EL_4, KS_2, CvM_2, AD_2, KS_3, CvM_3, and AD_3 for the bivariate normal distribution with correlation $\rho \in \{-0.5, -0.4, \ldots, 0.2\}$ and sample size $n = 100$. The significance level is 0.05 which is depicted by a dotted horizontal line.

pendence structures which are determined by Clayton and Frank copulas. These two copulas where also considered by Gijbels and Sznajder (2013). Each of them is controlled by a single parameter. Nelsen (2006) discussed that this parameter has a bijective and monotone relationship with Kendall's tau coefficient, denoted by τ. In other words, the two copulas can be described sufficiently by τ. Further, it can be shown that when C is a Clayton or Frank copula, X and Y are PQD if τ is nonnegative; are independent if $\tau = 0$; are not PQD if $\tau < 0$.

As in the bivariate normal examples, we choose Kendall's $\tau \in \{-0.9, -0.8, \ldots, 0.9\}$ for both copula families and use $1,000$ Monte Carlo replications to estimate rejection rates for each τ. We consider $n = 100$ and $\alpha = 0.05$. Figure 3.2 plots the curves of rejection rates along with τ taking values from -0.4 to 0.1. Results for τ outside this region were omitted because they are all zero when $\tau \geq 0.2$ and one when $\tau \leq -0.5$. Again, we see that all testing procedures are consistent and independence leads to the least favorable configuration for each method; i.e., powers goes to one when τ decreases to -0.4, sizes are around 0.05 when $\tau = 0$ and less than 0.05 when $\tau > 0$. Additionally, for the Clayton copula, all rejection rates are very close to each other except that the KS-type methods seems a little bit conservative. For the Frank copula, powers of EL_3 and the AD-type tests are close, powers of EL_4 and the CvM-type tests are close, and KS-type tests are still conservative. Lastly, we want to emphasize again that the EL_3 testing procedure has the largest powers in all settings.

3.5 Real Data Analysis

The dependence of assets is of key importance to portfolio construction and risk management (Elton and Gruber, 1973; Dempster, 2005). After the seminal work of Lehmann (1966), PQD has also been recognized as one important dependence for evaluating a combination of two financial assets (Denuit et al., 2005). For example, when two stocks are PQD, their combination would contribute limitedly to the risk reduction of a portfolio because the probability of their prices getting simultaneously low is higher than the one when they are independent. Thus, knowing the existence of PQD among stocks could help investors improve their investment strategy. In order to illustrate the use of our testing procedures for such applications, we consider three stocks: NASDAQ: APPL (Apple Inc.), NASDAQ: GOOGL (Google Inc.), and NYSE: WMT (Wal-Mart Stores, Inc.). Apple Inc. and Google Inc. are representative high technology companies, while Wal-Mart Stores Inc. is one of the leading companies in the retail industry.

We collected the closing prices of these three stocks (AAPL, GOOGL, WMT) from January 4th, 2016 to June 30th, 2016. In total, there are 125 trading days; i.e., $n = 125$. We plot the closing prices across each pair of stocks in the first row of Figure 3.3. For a clear and direct view of PQD, the second row of Figure 3.3 presents the scatter plots of the corresponding pseudo-observations defined in Gijbels

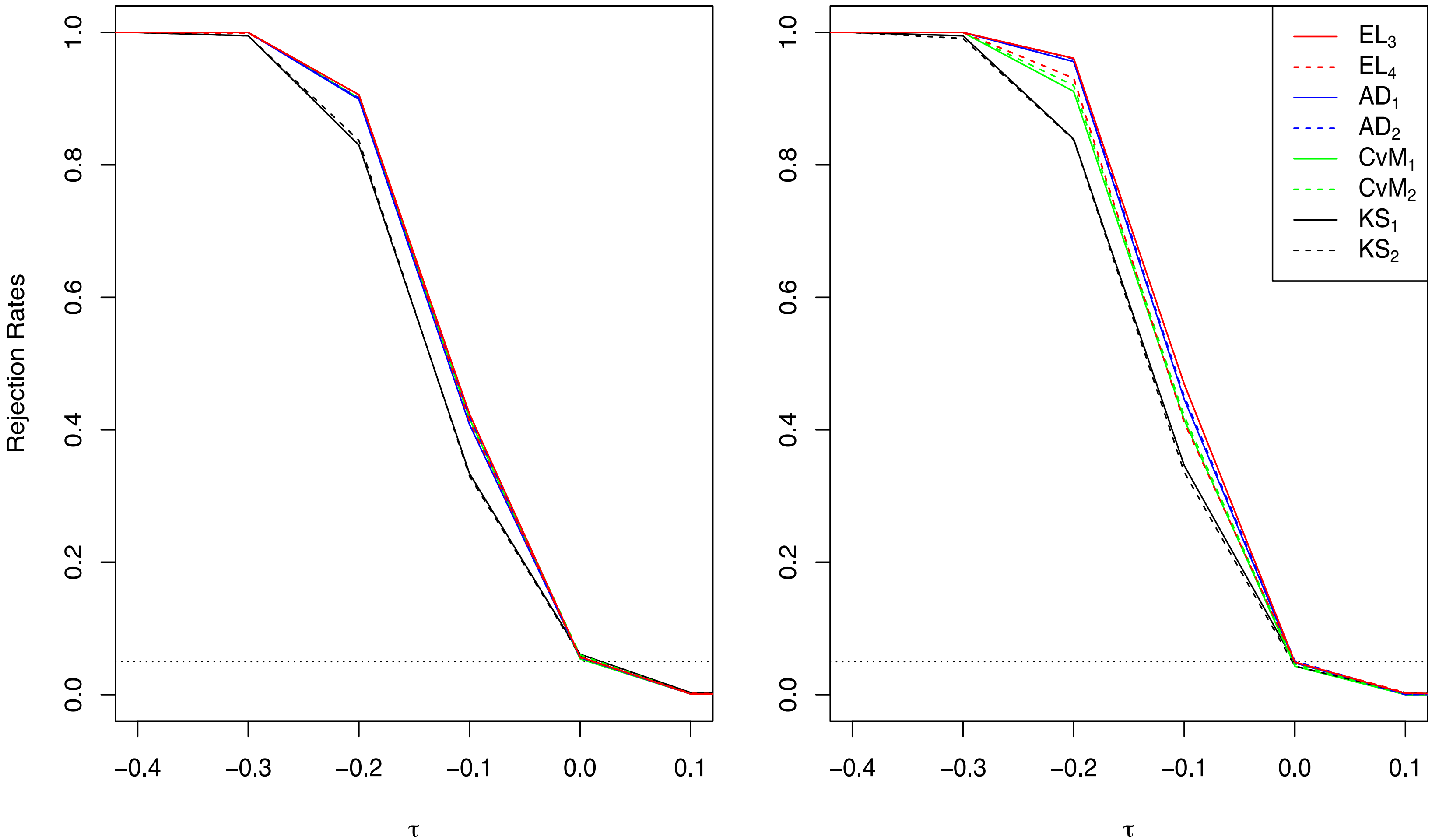

Figure 3.2: Left: Rejection rates of testing procedures EL_3, EL_4, KS_2, CvM_2, AD_2, KS_3, CvM_3, and AD_3 for Clayton copula (left) and Frank copula (right) with Kendall's $\tau \in \{-0.4, -0.4, \dots, 0.1\}$ and sample size $n = 100$. The significance level is 0.05 which is depicted by a dotted horizontal line.

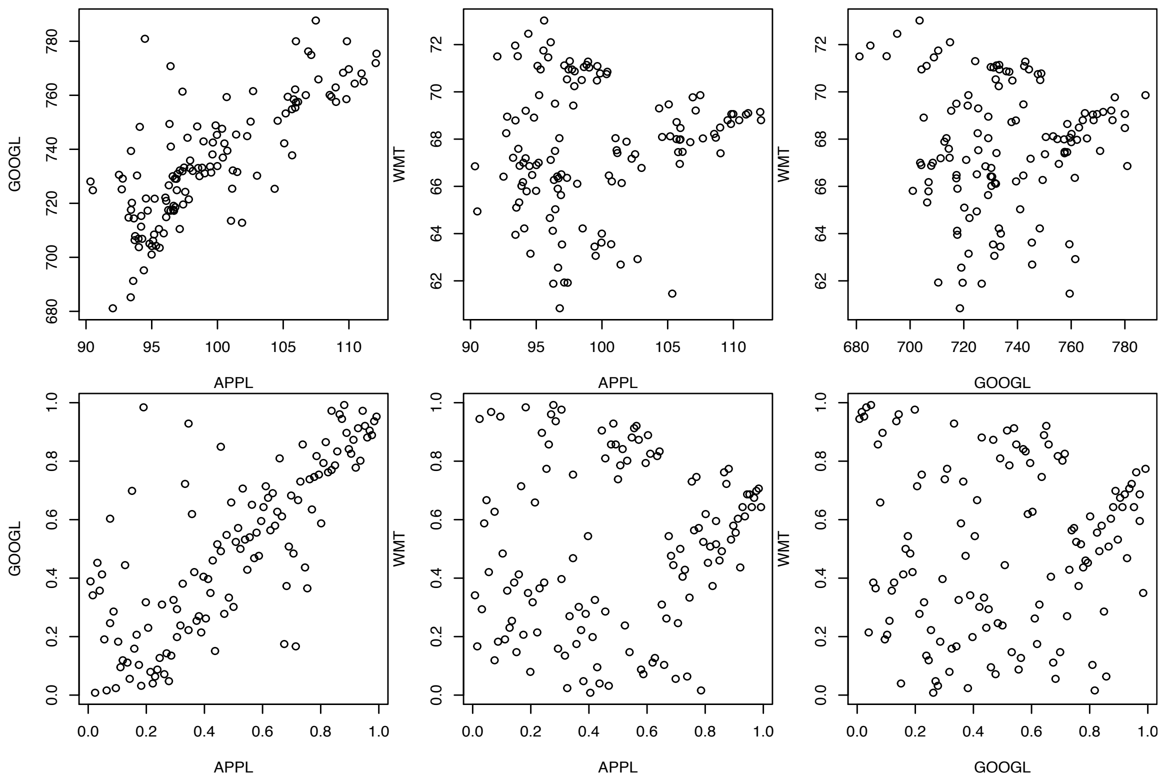

Figure 3.3: Scatter plots of stock prices (the first row) and the corresponding pseudo-observations (the second row). From left to right are APPL versus GOOGL, APPL versus WMT, and GOOGL versus WMT.

and Sznajder (2013). Generally speaking, suppose we have observations $\{(X_i, Y_i)\}_{i=1}^n$ of (X, Y); the pseudo-observations are $\{(U_i, V_i)\}_{i=1}^n$ where $U_i = nF_n(X_i)/(n+1)$ and $V_i = nG_n(Y_i)/(n+1)$. This transformation condenses the data relationship into the unit square (without changing possible PQD) and further eliminates the influence of extreme values. A scatter plot of the pseudo-observations would provide rough ideas about the dependence between X and Y. For example, if the locations of pseudo-observations are close to the diagonal line of the unit square, this suggests the existence of PQD; if pseudo-observations are spread uniformly within the unit square, X and Y are more likely independent. From Figure 3.3, one can see that PQD might exist between APPL and GOOGL, but not between APPL and WMT or GOOGL and WMT. To test these conjectures, because the scatter plots do not provide clear parametric forms for H or C, it is convenient and reasonable to consider

Table 3.3: The dependence coefficients, Pearson's rho, Kendall's tau, and Spearman's rho are provided for three stock prices among APPL, GOOGL, and WMT.

	(APPL, GOOGL)	(APPL, WMT)	(GOOGL, WMT)
Pearson's rho	0.773	0.109	-0.012
Kendall's tau	0.602	0.082	0.026
Spearman's tau	0.768	0.126	0.028

nonparametric testing approaches.

We first study APPL and GOOGL by testing the hypotheses $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$. The significance level throughout this section is chosen to be $\alpha = 0.05$. In addition to the scatter plots, the Pearson's rho, Kendall's tau, and Spearman's rho coefficients between APPL and GOOGL (listed in Table 3.5) all suggest that these two stocks are positively dependent. Thus, it is not surprising to see that all considered testing procedures reach the same conclusion; i.e., rejecting the null hypothesis $\mathcal{H}_0$. Specifically, the test statistics $\text{EL}_1 = 22.074$, $\text{EL}_2 = 34.083$, $\text{EM} = 22.081$ are larger than the estimated critical values 1.378, 1.711, and 1.818, respectively, and the distance-based test statistics $\text{KS}_1 = 2.088$, $\text{CvM}_1 = 1.469$, and $\text{AD}_1 = 40.241$ are larger than the estimated critical values 0.676, 0.070, and 3.446, respectively. However, not being independent does not imply the existence of PQD. Naturally, we proceed to consider the goodness-of-fit test: $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$. The results are $\text{EL}_3 = 0.007$, $\text{EL}_4 = 0.002$, $\text{KS}_2 = 0$, $\text{CvM}_2 = 0$, $\text{AD}_2 = 0$, which are less than critical values 1.355, 1.120, 0.546, 0.028, and 0.748, respectively. In addition, based on $1,000$ bootstrap samples, KS_3, CvM_3, and AD_3 each yield a p-value of 0.995. All eight testing procedures do not reject $\mathcal{H}_1$. In other words, these data do not provide sufficient evidence to reject PQD between the closing prices of APPL and GOOGL.

The existence of PQD between the prices of APPL and GOOGL might be easy to accept because both Apple Inc. and Google Inc. are successful high technology companies. However, Wal-Mart Stores Inc. is in the retail industry, so the type of dependence between APPL and WMT (named by Case one) and the one between GOOGL and WMT (named by Case two) might be different. From Figure 3.3, we see that the scatter plots of both cases (the middle and right columns of Figure 3.3) are more widely dispersed when compared to the ones between APPL and GOOGL (the left column of Figures 3.3). The Pearson's rho, Kendall's tau, and Spearman's tau coefficients in Table 3.5 are also much smaller than the ones between APPL and GOOGL. However, it is still unclear to us whether PQD holds or not.

To better reveal the dependence for these two cases, we first test $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$. For Case one, we have $\mathrm{EL}_1 = 2.814$, $\mathrm{EL}_2 = 3.292$, $\mathrm{EM} = 4.216$, $\mathrm{KS}_1 = 0.974$, $\mathrm{CvM}_1 = 0.126$, and $\mathrm{AD}_1 = 3.797$, which are greater than the corresponding critical values given earlier. All six testing procedures suggest rejecting the hypothesis that the closing prices of APPL and WMT are independent. For Case two, the test statistics are $\mathrm{EL}_1 = 1.481$, $\mathrm{EL}_2 = 1.740$, $\mathrm{EM} = 3.562$, $\mathrm{KS}_1 = 0.795$, $\mathrm{CvM}_1 = 0.077$, and $\mathrm{AD}_1 = 2.374$. Except for AD_1, all procedures reject independence between GOOGL and WMT. There is strong evidence that neither APPL and WMT nor GOOGL and WMT are independent.

We further test $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$. Interestingly, discrepancies now show up among distance-based testing procedures. For Case one, $\mathrm{KS}_2 = 0.541$, $\mathrm{CvM}_2 = 0.018$, and $\mathrm{AD}_2 = 0.678$ are all smaller than the estimated critical values 0.546, 0.028, and 0.748, respectively. With $1,000$ bootstrap samples, the KS_3, CvM_3, and AD_3 statistics provide estimated p-values 0.046, 0.095, and 0.057, respectively. One can see that, among distance-based methods, only KS_3 suggests rejecting the existence of PQD between APPL and WMT. For Case two, $\mathrm{KS}_2 = 0.427$ and $\mathrm{CvM}_2 = 0.020$ are both smaller than estimated critical values, but $\mathrm{AD}_2 = 1.131$ is too large and leads to rejecting $\mathcal{H}_1$. Further, the estimated p-values for KS_3, CvM_3, and AD_3, are 0.159, 0.094, and 0.014, respectively. Method AD_3 also rejects the null hypothesis. When such discrepancies occur, it is not clear which method(s) we should trust, and we might need to collect more data in order to make all distance-based methods reach the same conclusion. Such concerns do not arise for our EL-based tests, at least in this application. We have $\mathrm{EL}_3 = 1.402$ and $\mathrm{EL}_4 = 1.457$ for Case one and $\mathrm{EL}_3 = 2.081$ and $\mathrm{EL}_4 = 1.635$ for Case two. The estimated critical values for EL_3 and EL_4 are 1.355 and 1.120, respectively. Thus, with this sample of size $n = 125$, both EL-based testing procedures agree and suggest rejecting $\mathcal{H}_1$; i.e., neither APPL and WMT nor GOOGL and WMT are positive quadrant dependent.

3.6 Conclusion

In this chapter, we have proposed EL-based testing procedures to test for and against PQD. Our tests are more straightforward than parametrization-based and distance-based testing procedures. More detailed approximations are needed to provide rigorous proofs for Conjecture 3.1. However, numerical evidence not only supports the theoretical results but also suggests that our testing procedures provide better power. In the data analysis, we reach the conclusion that APPL versus WMT and GOOGL versus WMT are both not PQD. The distance-based testing procedures provide mixed results.

From the stock price data, we have considered the pairwise existence of PQD among stock prices; i.e., the probability two stock prices provide low values simultaneously is higher than it would be if they are independent. This idea can be naturally extended to consider three or more stocks which is called positive orthant dependence. Generalizing our EL-bsed tests to higher dimensions is left for future work. We expect EL-based procedures would be more straightforward than the same tests based on functional distance and these based on reparametrization.

Chapter 4

Extensions

In Chapter 2, we proposed ODC-based goodness-of-fit tests for uniform stochastic ordering (USO) when the data are not censored. In Chapter 3, we proposed EL-based testing procedures to investigate positive quadrant dependence. Both methods provide attractive advantages over existing work. In this chapter, we follow the route to discuss possible extensions of each of the two methods. We first generalize the ODC-based testing procedure to incorporate right-censored time-to-event data in Section 4.1. We then build EL-based testing procedures to test for exchangeability in Section 4.2. Only the main ideas are provided.

4.1 Nonparametric goodness-of-fit tests for uniform stochastic ordering (USO) with random right-censored data

Summary: In Chapter 2, we proposed ODC-based goodness-of-fit tests for uniform stochastic ordering with complete data. If observations are time-to-event data, they are often randomly right-censored. Thus, it is natural to extend our ODC-based methods to incorporate such observations. The ODC-based method we developed in Chapter 2 consists of three parts: constructing an estimator of the ODC, calculating its least star-shaped majorant, and determining rejection regions. When observations are randomly right-censored, one can easily construct an estimator of the ODC curve by using Kaplan-Meier estimators and then following Chapter 2 to calculate the least star-shaped majorant of the estimator. However, finding rejection regions is challenging because of the unknown censoring distribution. In this section, we focus

on two cases and suggest modified testing procedures.

4.1.1 Introduction

Stochastic orderings are used to order random variables by comparing their distribution functions. Consider two random variables X and Y with distribution functions F and G, respectively and survival functions $\overline{F} = 1 - F$ and $\overline{G} = 1 - G$. With censored data being commonly seen in survival analysis, it is natural to extend our ODC-based tests proposed in Chapter 2 to incorporate such observations. However, difficulties arise in finding the least favorable configuration of F and G when censoring is involved. In Section 4.1.2, we illustrate the construction of new testing procedures that incorperate censoring and present asymptotic distributions of suitably modified test statistics.

4.1.2 Testing procedures and asymptotic distributions of test statistics

Let $T_{11}, \ldots, T_{1m}$ and $T_{21}, \ldots, T_{2n}$ be independent random samples from lifetime distributions F and G, respectively. Our goal is to test $H_0 : F \leq_{\text{US}} G$ versus $H_1 : F \not\leq_{\text{US}} G$. Denote by U_{ij} the censoring time associated with T_{ij}. We assume that $U_{11}, \ldots, U_{1m}$ and $U_{21}, \ldots, U_{2n}$ are independent random variables from distributions L_1 and L_2, respectively. Our observed data are $X_{ij} = \min\{T_{ij}, U_{ij}\}$ and the censoring indicators $\delta_{ij} = I(T_{ij} \leq U_{ij})$ for $i = 1(2)$ and $j = 1, \ldots, m(n)$, where I is the indicator function. If $\delta_{ij} = 1$, we observe $X_{ij} = T_{ij}$; otherwise, T_{ij} is censored and we have U_{ij}.

G is known:

We start with a simpler case where G is assumed to be known. Consider the transformed random variables $\widetilde{T}_j = G(T_{1j})$, $\widetilde{U}_j = G(U_{1j})$, $\widetilde{X}_j = \min\{\widetilde{T}_j, \widetilde{U}_j\}$, and $\delta_j = I(\widetilde{T}_j \leq \widetilde{U}_j)$ for $j = 1, \ldots, m$. It can be shown that $\{\widetilde{T}_j\}_{j=1}^m$ is a random sample from the distribution $R = FG^{-1}$. In other words, $\{\widetilde{X}_j\}_{j=1}^m$ is sufficient to estimate the ODC. A Kaplan-Meier-based estimator $\widehat{R}_m$ of R is given by

$$\widehat{R}_m(u) = 1 - \prod_{v \le u} \left\{ 1 - \frac{\Delta \overline{N}_1^G(v)}{\overline{Y}_1^G(v)} \right\},$$

where $\Delta h(v) = h(v) - \lim_{\delta \downarrow 0} h(v-\delta)$, $\overline{N}_1^G(v) = \sum_{j=1}^m I(\widetilde{X}_j \le v, \delta_j = 1)$ is a counting process, and $\overline{Y}_1^G(v) = \sum_{j=1}^m I(\widetilde{X}_j \ge v)$ is the at-risk process.

According to the asymptotic distribution of Kaplan-Meier estimators, we obtain the limiting distribution of $\widehat{R}_m$.

Lemma 4.1. *Assume R is continuous, G is known, and suppose $\widetilde{U}_j \perp\!\!\!\perp \widetilde{T}_j$ for all j. Then,*

$$\sqrt{m}\{\widehat{R}_m(u) - R(u)\} \xrightarrow{L} \{1 - R(u)\}\mathcal{W}_1\{\nu_1(u)\}, \text{ for all } u \in [0, t_1],$$

as $m \to \infty$, where $\xrightarrow{L}$ means convergence in law, $\mathcal{W}_1$ is a standard Wiener process, $\nu_1(u) = \int_0^u 1/\pi_1(v)\, d\Lambda_1(v)$, Λ_1 is the cumulative hazard rate function of F, $\pi_1(u) = \mathrm{pr}(\widetilde{X}_j \ge u)$, and $t_1 = \sup\{u : \pi_1(u) > 0\}$.

From Lemma 4.1, we see that $\widehat{R}_m$ is a consistent estimator of R. It is important to note that the censoring distribution affects the limiting distribution through ν_1, which is a quantity depending on the censoring distribution L_1.

To construct test statistics, we follow the idea in Chapter 2. Let $\mathcal{M}$ be the least star-shaped majorant operator mapping from $l([0,1])$ to $l([0,1])$. Our restricted estimator of R under USO is $\mathcal{M}\widehat{R}_m$. According to the Lipschitz continuity property of $\mathcal{M}$ provided in Lemma A.3, $\mathcal{M}\widehat{R}_m$ is also a consistent estimator of R. Further, it is clear that the censoring distribution affects the limiting distribution of $\mathcal{M}\widehat{R}_m$ through ν_1 as well.

To explore the limiting distribution $\mathcal{M}\widehat{R}_m$ in more detail, we again need to bifurcate $[0,1]$ into two regions: non-strictly star-shaped regions $\cup_k[a_k, b_k]$ and the strictly star-shaped region $[0,1]\backslash \cup_k [a_k, b_k]$. We refer readers to Section 2.3.1 for a detailed construction of these two regions. Because the limiting distribution of $\mathcal{M}\widehat{R}_m$ depends on the censoring distribution through ν_1, we first assume that ν_1 is known and consider a modified process

$$\sqrt{m}\left[\frac{\mathcal{M}\widehat{R}_m(\nu_1^{-1}(u)) - \widehat{R}_m(\nu_1^{-1}(u))}{\{1-\widehat{R}_m(\nu_1^{-1}(u))\}^{1/2}}\right] \quad \text{for } u \in [0, t_1].$$

Given a fixed u in a non-strictly star-shaped region $[a_k, b_k]$, according to the delta method, we have the following pointwise asymptotic distribution; i.e.,

$$\sqrt{m}\left[\frac{\mathcal{M}\widehat{R}_m(\nu_1^{-1}(u)) - \widehat{R}_m(\nu_1^{-1}(u))}{\{1-\widehat{R}_m(\nu_1^{-1}(u))\}^{1/2}}\right] \xrightarrow{d} \{\mathcal{M}^{(1,0)}_{[a_k,b_k]}\mathcal{B}_1\}(u),$$

where $\mathcal{M}^{(1,0)}_{[a_k,b_k]}$ is defined in Section 2.3.1 and $\mathcal{B}_1$ is a standard Brownian bridge. If u is in the strictly star-shaped region, then the modified process at u converges to zero. Note that the limiting distributions do not depend on ν_1. In other words, if ν_1 is known, we can get rid of the effect caused by ν_1.

When ν_1 is unknown, a well-known consistent estimator of ν_1 is given by

$$\widehat{\nu}_1(u) = m\int_0^u [\{\overline{Y}_1^G(v) - \Delta\overline{N}_1^G(v)\}\overline{Y}_1^G(v)]^{-1}\, d\overline{N}_1^G(v).$$

Consequently, we can estimate ν_1^{-1} by $\widehat{\nu}_1^{-1}(u) = \inf_{t\geq 0}\{\widehat{\nu}_1(t) \geq u\}$. Now, we are able to define our L^p distance-based test statistics adjusted by censoring; i.e., for all $p \in [1, \infty)$,

$$M_m^p = \sqrt{m}\left\{\int_0^{t_1}\left(\frac{\mathcal{M}\widehat{R}_m(\widehat{\nu}_1^{-1}(u)) - \widehat{R}_m(\widehat{\nu}_1^{-1}(u))}{[1-\widehat{R}_m(\widehat{\nu}_1^{-1}(u))]^{1/2}}\right)^p du\right\}^{1/p}$$

and

$$M_m^\infty = \sqrt{m}\sup_{u\in[0,t_1]}\frac{\mathcal{M}\widehat{R}_m(\widehat{\nu}_1^{-1}(u)) - \widehat{R}_m(\widehat{\nu}_1^{-1}(u))}{[1-\widehat{R}_m(\widehat{\nu}_1^{-1}(u))]^{1/2}}.$$

Large values of M_m^p, for $p \in [1, \infty]$, are evidence against H_0. Using Lemma A.6, we should be able to obtain the asymptotic distribution of our test statistics. However, the limiting behavior of $\mathcal{M}\widehat{R}_m(\widehat{\nu}_1^{-1}(u))$ is more complicated because of the estimated inverse function $\widehat{\nu}_1^{-1}$. So we state our conjectures below.

Conjecture 4.1. Following the same assumptions stated in Lemma 4.1, when H_0 is true, we have

$$M_m^p \xrightarrow{d} \left(\sum_k \int_{[0,\nu_1^{-1}(t_1)]\cap[a_k,b_k]} [\mathcal{M}_{[a_k,b_k]}^{(0,1)} \mathcal{B}(u)]^p \, du \right)^{1/p}$$

$$M_m^\infty \xrightarrow{d} \sup_k \sup_{u\in[0,\nu_1^{-1}(t_1)]\cap[a_k,b_k]} \mathcal{M}_{[a_k,b_k]}^{(0,1)} \mathcal{B}(u),$$

as $n \to \infty$, where $\mathcal{B}$ is a standard Brownian bridge.

Following the same discussion as in the proof of Theorem 2.2, one can see that $R(u) = u$, or equivalently, $F = G$, yields the least favorable configuration.

Conjecture 4.2. Following the same assumptions stated in Lemma 4.1, when H_0 is true, we have that

$$\mathrm{pr}(M_m^p > c) \leq \mathrm{pr} \left(\left\{ \int_{[0,\nu_1^{-1}(t_1)]} [\mathcal{M}_{[0,1]}^{(0,1)} \mathcal{B}(u)]^p \, du \right\}^{1/p} > c \right)$$

$$\mathrm{pr}(M_m^\infty > c) \leq \mathrm{pr} \left(\sup_{u\in[0,\nu_1^{-1}(t_1)]} \mathcal{M}_{[0,1]}^{(0,1)} \mathcal{B}(u) > c \right)$$

holds for any $c > 0$.

Note that the interval $[0, \nu_1^{-1}(t_1)]$ is contained in $[0, 1]$. Thus, regardless of the value of $\nu_1^{-1}(t_1)$, the same critical values in Table A.1 can be used to construct a conservative test. If $\nu_1^{-1}(t_1)$ can be estimated, then refined critical values can be estimated according to the limiting distribution in Conjecture 4.2.

G IS UNKNOWN

When G is unknown, we first assume that the lifetime variables T_{2j}'s from G are not censored. Then we can use the empirical distribution G_n to estimate G and use the Kaplan-Meier estimator $\widehat{F}_m$ to estimate F. Consequently, we have an estimator of R by $\widehat{R}_{mn} = \widehat{F}_m G_n^{-1}$ where G_n^{-1} is the empirical quantile function associated with G_n. The following lemma states the asymptotic distribution of $\widehat{R}_{mn}$.

Lemma 4.2. *Follow the same assumptions stated in Lemma 4.1, we further assume that the first derivative R' of R is continuous and bounded over* $(0,1)$. *When H_0 is true,* $\min\{m,n\} \to \infty$, *and* $n/(m+n) \to \lambda \in (0,1)$, *we have*

$$c_{mn}\{\widehat{R}_{mn}(u) - R(u)\} \xrightarrow{L} \lambda^{1/2}\{1-R(u)\}\mathcal{W}_1\{\nu_1(u)\} + (1-\lambda)^{1/2}R'(u)(1-u)\mathcal{W}_2(u),$$

for all $u \in [0,t_2]$, where $c_{mn} = (mn/(m+n))^{1/2}$, $t_2 = \sup\{u : \pi_2(u) > 0\}$, $\pi_2(u) =$ $\mathrm{pr}(\min\{X_{1j}, X_{2j}\} \geq u)$, *and $\{\mathcal{W}_i\}_{i=1}^2$ are two independent Wiener processes.*

We further simplify the above limiting distribution to see if ν_1 can be canceled; i.e., if the censoring distribution does not affect the result. Under H_0, note that $1 - R(u) = R'(u)(1-u)$ when u is in a non-strictly star-shaped region; then, we can rewrite the limiting process within $[a_k, b_k]$ as

$$c_{mn}\{\widehat{R}_{mn}(u) - R(u)\} \xrightarrow{L} \{1-R(u)\}[\lambda^{1/2}\mathcal{W}_1\{\nu_1(u)\} + (1-\lambda)^{1/2}\mathcal{W}_2(u)].$$

Note that the limiting distribution contains $\mathcal{W}_1\{\nu_1(u)\}$ and $\mathcal{W}_2(u)$ simultaneously. Thus, in this case, ν_1 cannot be removed. This finding brings difficulty to establish the least favorable configuration. More studies have to done in this case.

4.2 Empirical-likelihood-based testing for exchangeability

Summary: The dependence structure between two random variables is not limited to linear dependence or positive quadrant dependence. Exchangeability is also one of the most important dependence structures in statistics. Testing for exchangeability can be used to verify whether there exists a difference between two dependent treatments. Motivated by tests proposed in Chapter 3, we now build EL-based testing procedures to test for exchangeability. The critical values are selected from a permutation method. We illustrate the method using clinical trail data, which was examined by Hollander (1971).

4.2.1 Introduction

Consider a bivariate random vector (X, Y) which follows a joint distribution function H. A random vector (X, Y) is exchangeable simply means that $H(x, y) = H(y, x)$ for all x and y. In other words, the joint distribution function H is symmetric with respect to the diagonal line $y = x$. Exchangeability of (X, Y) is also called interchangeability or symmetry (Sen, 1967; Hollander, 1971). It has been used in many applications. For example, when an experimental unit is measured twice in the control and treatment groups at the same time, the intrinsic dependence structure between repeated measurements naturally exists. It is not realistic to assume independence to compare between distributions under control and treatment. Instead, researchers turn to test for exchangeability to see if there is an effect due to treatment (Bell and Haller, 1969).

To test for exchangeability, Hollander (1971) proposed a conditional distribution-free testing procedure by estimating a departure of the joint distribution from being exchangeable. Utilizing the minimum spanning tree introduced by Friedman and Rafsky (1979), Modarres (2008) generalized the one-dimensional Wald-Wolfowitz runs test and the Smirnov rank test to a two-dimensional runs test, rank test, and nearest neighbor test. Conservative powers are expected because the original one-dimensional runs test and rank test are already conservative.

Instead of developing fully nonparametric testing procedures, many authors have focused on specific distribution families. Consider a bivariate distribution family with marginals controlled by a location and a scale parameter. Because exchangeability implies identical location and scale parameters, Sen (1967) proposed a rank test to test against unequal location or scale parameters. Kepner and Randles (1982) proposed a more powerful distribution-free testing procedure than Sen (1967) to test against unequal marginal distributions with different scale parameters. Kepner and Randles (1984) further extended their work to test against unequal marginal distributions with different location or scale parameters. Instead of considering a bivariate distribution family, Elton and Gruber (1999) parametrized the hypotheses with means and a variance-covariance matrix and then proposed testing procedures against unequal marginal distributions with means or variances. These approaches focused on smaller parametrized alternative hypotheses. However, it might also have a loss of power when alternatives are not recruited in the considered distribution family or when the first or second moments do not exist.

Without restricting to a specific distribution family, Yanagimoto and Sibuya (1976) focused on alternative hypotheses which can be described by proper stochastic orderings (Yanagimoto and Sibuya, 1972) and proposed one-sided conditional sign tests. Snijders (1981) further considered one-sided rank tests based on a concept of "asymmetry towards high X-value" which is equivalent to particular stochastic orderings under alternatives. However, it is clear that if an alternative cannot be described equivalently by the stochastic orderings they have considered, these approaches might be inadequate.

In this chapter, we consider a fully nonparametric approach; i.e., we do not focus on a specific distribution family or modify the null or the alternative hypotheses. Our approach follows the same empirical likelihood method used in Chapter 3.

4.2.2 Testing procedure

Suppose we have a random sample $\{(X_i, Y_i)\}_{i=1}^n$ from the distribution H. Our goal is to test $\mathcal{H}_0$ versus $\mathcal{H}_1$, where

$$\begin{cases} \mathcal{H}_0: & H(x,y) = H(y,x) \text{ for all } x, y \in \mathbb{R} \\ \mathcal{H}_1: & H(x,y) \neq H(y,x) \text{ for some } x, y \in \mathbb{R}. \end{cases}$$

Similar to the testing procedures proposed in Chapter 3, we first localize the hypotheses to obtain local statistics, and we then aggregate the local test statistics to obtain a global test statistic. Fixing $(x, y)' \in \mathbb{R}^2$, consider the local hypotheses $\mathcal{H}_0(x, y)$ and

$\mathcal{H}_1(x,y)$, where

$$\begin{cases} \mathcal{H}_0(x,y): & H(x,y) = H(y,x) \\ \mathcal{H}_1(x,y): & H(x,y) \neq H(y,x). \end{cases}$$

The local empirical likelihood ratio $R(x,y)$ can be obtained by

$$\begin{aligned} R(x,y) &= \frac{\sup\{L(\tilde{H}) : \tilde{H}(x,y) = \tilde{H}(y,x)\}}{\sup\{L(\tilde{H})\}} \\ &= \left(\frac{P_n(A_v) + P_n(A_h)}{2P_n(A_v)}\right)^{nP_n(A_v)} \left(\frac{P_n(A_v) + P_n(A_h)}{2P_n(A_h)}\right)^{nP_n(A_h)}, \end{aligned}$$

where

$$\begin{aligned} A_h &= (-\infty, \min\{x,y\}] \times [\min\{x,y\}, \max\{x,y\}) \\ A_v &= [\min\{x,y\}, \max\{x,y\}) \times (-\infty, \min\{x,y\}]. \end{aligned}$$

We refer readers to Chapter 3 for the definition of $L(\cdot)$. Appendix C provides detailed derivations of the supremum of empirical likelihood functions.

Similar to most EL-based testing procedures, we choose $-2\ln R(x,y)$ as a local test statistic. According to the second-order Taylor expansion of $\ln(1+a)$ at $a=0$ and the delta method, one can show that $-2\ln R(x,y)$ converges to Z where Z follows the standard normal distribution. We can see that the local test statistics are automatically self-standardized. This reveals a possible advantage of the EL-based methods.

Now we construct our global test statistic T_n by integrating $-2\ln R(x,y)$ with respect to the joint empirical distribution function H_n; i.e.,

$$T_n = \int_{(x,y)\in\mathbb{R}^2} -2\ln R(x,y)\, dH_n(x,y) = -\frac{2}{n}\sum_{i=1}^{n} \ln R(X_i, Y_i).$$

For convenience, we define $\ln\infty = 0$. Large values of T_n are evidence against $\mathcal{H}_0$.

4.2.3 Asymptotic distribution of T_n

The limiting distribution of a local test statistic is clear. Unfortunately, establishing the asymptotic distribution of the global test statistic T_n is not. We state our finding as a conjecture.

Conjecture 4.3. Assume both X and Y are continuous. Under $\mathcal{H}_0$; i.e., $H(x, y) = H(y, x)$ for all x, y, we have

$$T_n \xrightarrow{d} \int_{\mathbb{R}^2} \frac{[\mathbb{G}(x, y) - \mathbb{G}(y, x)]^2}{H(x, y) - H(\min\{x, y\}, \min\{x, y\})} \, dH(x, y)$$

where $\mathbb{G}$ is a mean zero Gaussian process with covariance

$$Cov(\mathbb{G}(x_1, y_1), \mathbb{G}(x_2, y_2)) = H(\min\{x_1, x_2\}, \min\{y_1, y_2\}) - H(x_1, y_1)H(x_2, y_2).$$

Remark 4.1. Under $\mathcal{H}_0$, the limiting behavior of the integral is not clear because the denominator $H(x, y) - H(\min\{x, y\}, \min\{x, y\})$ is not necessary bounded away from zero over the support of H. Necessary assumptions are needed.

4.2.4 Numerical approaches

To illustrate the performance of our testing procedure, we need to determine the critical value such that Type I error probabilities are controlled. The solution we come up with is to use a permutation method. This method has also been used by Modarres (2008).

Permutation method

For any distribution function H, it is easy to construct a distribution H_E such that $H_E(x, y) = H_E(y, x)$ for all $x, y \in \mathbb{R}$. This distribution can be obtained by

$$H_E(x, y) = \frac{H(x, y) + H(y, x)}{2}.$$

When X and Y are truly exchangeable, then $H_E = H$. The function H_E is easy to construct, and it always satisfies the null hypothesis. Thus, we can estimate critical values from H_E.

We introduce a random variable ξ which is independent of (X, Y) and follows Bernoulli(0.5); i.e., Bernoulli distribution with success probability 0.5. Then we consider a new bivariate random vector (X^*, Y^*) where

$$(X^*, Y^*) = \begin{cases} (X, Y), & \text{if } \xi = 0 \\ (Y, X), & \text{if } \xi = 1. \end{cases}$$

It is easy to see that $(X^*, Y^*) \sim H_E$. In other words, by independently flipping (X, Y) into (Y, X) with probability 0.5, a newly generated random vector (X^*, Y^*) follows H_E. Using (X^*, Y^*), we can appropriately determine rejection regions. The following steps describe our testing procedure:

1. Compute the test statistic T_n from the observation $\{(X_i, Y_i)\}_{i=1}^n$.

2. For each (X_i, Y_i), $i = 1, \ldots, n$, randomly flip the position of X_i and Y_i; i.e., draw ξ_i from Bernoulli(0.5) to get

$$(X_i^*, Y_i^*) = (1 - \xi_i)(X_i, Y_i) + \xi_i(Y_i, X_i).$$

3. Calculate T_n^* from $\{(X_i^*, Y_i^*)\}_{i=1}^n$.

4. Repeat Steps 2 and 3 for B times, where B is large. Denote the obtained test statistics by $T_n^{*(1)}, \ldots, T_n^{*(B)}$.

5. Approximate the distribution of T_n using $\{T_n^{*(k)}\}_{k=1}^B$. The critical value c_α can be estimated by using the upper $1 - \alpha$ sample quantile of $\{T_n^{*(k)}\}_{k=1}^B$.

6. Reject $\mathcal{H}_0$ if $T_n > c_\alpha$.

4.2.5 Real data

Table 4.1 presents data used in Hollander (1971) from a double-blind clinical trial involving nine patients who were diagnosed as having mixed anxiety and depression. A tranquilizer was chosen to be the treatment. To measure each patient's level of anxiety and depression, a Hamilton depression scale factor IV was used. Lower values indicate lower levels of anxiety and depression. The data are values for each patient after their first (X) and second (Y) therapies. We want to test whether the tranquilizer truly made a difference. To answer this question, it is natural to test for the exchangeability between X and Y.

Applying our testing procedure introduced in the last section, we chose $B = 1,000$ and considered $\alpha = 0.05$. The critical c_α was estimated to be 0.555 which is smaller than test statistic $T_n = 0.672$. Hence, we reject the null hypothesis $\mathcal{H}_0$. In other words, the data provide strong enough evidence that X and Y are not exchangeable.

Table 4.1: Hamilton depression scale factor IV values on the tranquilizer, provided in Hollander (1971).

Patient i	1	2	3	4	5	6	7	8	9
First visit	1.83	0.50	1.62	2.48	1.68	1.88	1.55	3.06	1.30
Second visit	0.878	0.647	0.598	2.05	1.06	1.29	1.06	3.14	1.29

Bibliography

Aranda, J., Beharry, K., Valencia, G. Natarajan, G., and Davis, J. (2010). "Caffeine impact on neonatal morbidities". *Journal of Maternal-Fetal and Neonatal Medicine* 23, pp. 20–23.

Arcones, M. and Samaniego, F. (2000). "On the asymptotic distribution theory of a class of consistent estimators of a distribution satisfying a uniform stochastic ordering constraint". *Annals of Statistics* 28, pp. 116–150.

Bamber, D. (1975). "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph". *Journal of Mathematical Psychology* 12, pp. 387–415.

Beare, B. and Moon, J. (2015). "Nonparametric tests of density ratio ordering". *Econometric Theory* 31, pp. 471–492.

Bell, C. and Haller, H (1969). "Bivariate symmetry tests: Parametric and nonparametric". *Annals of Mathematical Statistics* 40, pp. 259–269.

Billingsley, P. (1999). *Convergence of Probability Measures*. New York: Wiley.

Carolan, C. and Tebbs, J. (2005). "Nonparametric tests for and against likelihood ratio ordering in the two-sample problem". *Biometrika* 92, pp. 159–171.

Cox, C., Hashem, N., Tebbs, J., Bookstaver, B., and Iskersky, V. (2015). "Evaluation of caffeine and the development of necrotizing enterocolitis". *Journal of Neonatal-Perinatal Medicine* 8, pp. 339–347.

Dardanoni, V. and Forcina, A. (1998). "A unified approach to likelihood inference on stochastic orderings in a nonparametric context". *Journal of the American Statistical Association* 93, pp. 1112–1123.

Davidov, O. and Herman, A. (2012). "Ordinal dominance curve based inference for stochastically ordered distributions". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74, pp. 825–847.

Deheuvels, P. (1979). “La fonction de dépendance empirique et ses propriétés.” *Academie Royale de Belgique, Bulletin de la Classe des Sciences, 5e Série* 65, pp. 274–292.

Dempster, M. (2005). *Risk Management: Value at Risk and Beyond 1st Edition*. Cambridge University Press.

Denuit, M. and Scaille, O. (2004). “Nonparametric tests for positive quadrant dependence”. *Journal of Financial Econometrics* 2, pp. 422–450.

Denuit, M., Dhaene, J., and Ribas, C. (2001). “Does positive dependence between individual risks increase stop-loss premiums?” *Insurance: Mathematics and Economics* 28, pp. 305–308.

Dhaene, J. and Goovaerts, M. (1996). “Dependency of risks and stop-loss order”. *ASTIN Bulletin* 26, pp. 201–212.

Dobson, N. and Hunt, C. (2013). “Caffeine use in neonates: Indications, pharmacokinetics, clinical effects, outcomes”. *NeoReviews* 14, pp. 540–550.

Drouet, D. and Kotz, S. (2001). *Correlation and Dependence*. London: Imperial College Press.

Dykstra, R., Kochar, S., and Robertson, T. (1991). “Statistical inference for uniform stochastic ordering in several populations”. *Annals of Statistics* 19, pp. 870–888.

Einmahl, J. and McKeague, I. (2003). “Empirical likelihood based hypothesis testing”. *Bernoulli* 9, pp. 267–290.

El Barmi, H. and McKeague, I. (2005). “Inferences under a stochastic ordering constraint: The k-sample case”. *Journal of the American Statistical Association* 100, pp. 252–261.

El Barmi, H. and McKeague, I. (2013). “Empirical likelihood-based tests for stochastic ordering”. *Bernoulli* 19, pp. 295–307.

El Barmi, H. and Mukerjee, H. (2016). “Consistent estimation of survival functions under uniform statistic ordering: The k-sample case”. *Journal of Multivariate Analysis* 144, pp. 99–109.

Elton, E. and Gruber, M. (1973). “Estimating the dependence structure of share prices-implications for portfolio selection”. *The Journal of Finance* 28, 1203Ð1232.

Elton, E. and Gruber, M. (1999). “A class of permutation tests of bivariate interchangeability”. *Journal of the American Statistical Association* 94, pp. 273–284.

Embrechts, P., McNeil, A., and Straumann, D. (2002). “Correlation and dependence in risk management: properties and pitfalls”. *Risk Management: Value at Risk and Beyond*. Cambridge: Cambridge University Press.

Friedman, J. and Rafsky, L. (1979). “Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests”. *Annals of Statistics* 2, pp. 697–717.

Gijbels, I. and Sznajder, D. (2013). “Positive quadrant dependence testing and constrained copula estimation”. *Canadian Journal of Statistics* 41, pp. 36–64.

Gijbels, I., Amelia, M., and Sznajder, D. (2010). “Positive quadrant dependence tests for copulas”. *Canadian Journal of Statistics* 38, pp. 555–581.

Hollander, M. (1971). “A nonparametric test for bivariate symmetry”. *Biometrika* 58, pp. 203–212.

Hsieh, F. and Turnbull, B. (1996). “Nonparametric estimation of the receiver operating characteristic curve”. *Annals of Statistics* 24, pp. 25–40.

Janic-Wròbelwska, A., Kallenberg, W.C.M., and Ledwina, T. (2004). “Detecting positive quadrant dependence and positive function dependence.” *Insurance: Mathematics and Economics* 34, pp. 467–487.

Kepner, J and Randles, R. (1982). “Detecting unequal marginal scales in a bivariate population”. *Journal of the American Statistical Association* 77, pp. 475–482.

Kepner, J and Randles, R. (1984). “Comparison of tests for bivariate symmetry versus location and or scale alternatives”. *Communications in Statistics - Theory and Methods* 13, pp. 915–930.

Kochar, S. and Gupta, R. (1987). “Competitors of the Kendall-tau test for testing independence against positive quadrant dependence”. *Biometrika* 74, pp. 664–666.

Komlós, J., Major, P., and Tusnády, G. (1975). “An approximation of partial sums of independent RV’s and the sample DF. I”. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 32, pp. 111–131.

Lai, C. (2003). “Concepts of stochastic dependence in reliability analysis”. *Handbook of Reliability Engineering*, pp. 141–156.

Lai, C. and Xie, M. (2006). *Stochastic Aging and Dependence for Reliability*. New York, NY: Springer Science and Business Media.

Ledwina, T. and Wyłupek, G. (2014). “Validation of positive quadrant dependence”. *Insurance: Mathematics and Economics* 56, pp. 38–47.

Lehmann, E. (1966). "Some concepts of dependence". *Annals of Mathematical Statistics* 37, pp. 1137–1153.

Lehmann, E. and Rojo, J. (1992). "Invariant directional orderings". *Annals of Statistics* 20, pp. 2100–2110.

Levy, H. (1992). "Stochastic dominance and expected utility: Survey and analysis". *Management Science* 38, pp. 555–593.

Modarres, R. (2008). "Tests of bivariate exchangeability". *International Statistical Review* 76, pp. 203–213.

Mukerjee, H. (1996). "Estimation of survival functions under uniform stochastic ordering". *Journal of the American Statistical Association* 91, pp. 1684–1689.

Nelsen, R. (2006). *An Introduction to Copulas, Lecture Notes in Statistics*. New York: Lecture Notes in Statistics, Springer.

Omelka, M., Gijbels, I., and Veraverbeke, N. (2009). "Improved kernel estimation of copulas: Week convergence and goodness-of-fit testing". *Annals of Statistics* 37, pp. 3023–3058.

Owen, A. (1990). "Empirical likelihood ratio confidence regions". *Annals of Statistics* 18, pp. 90–120.

Park, C., Lee, C., and Robertson, T. (1998). "Goodness-of-fit test for uniform stochastic ordering among several populations". *Canadian Journal of Statistics* 26, pp. 69–81.

Rojo, J. (2004). "On the estimation of survival functions under a stochastic order constraint". *Lecture Notes-Monograph Series* 44, pp. 37–61.

Rojo, J. and Samaniego, F. (1993). "On estimating a survival curve subject to a uniform stochastic ordering constraint". *Journal of the American Statistical Association* 88, pp. 566–572.

Scaillet, O. (2005). "A Kolmogorov-Smirnov type test for positive quadrant dependence". *Canadian Journal of Statistics* 33, pp. 415–427.

Schmidt, B., Roberts, R., Davis, P., Doyle, L., Barrington, K., Ohlsson, A., Solimano, A., and Tin, W. (2006). "Caffeine therapy for apnea of prematurity". *New England Journal of Medicine* 354, pp. 2112–2121.

Sen, P. (1967). “Nonparametric tests for multivariate interchangeability. Part 1: Problems of location and scale in bivariate distributions”. *Sankhyā: The Indian Journal of Statistics, Series A* 29, 351Ð372.

Shaked, M. and Shanthikumar, J. (1994). *Stochastic Orders and their Application.* New York: Academic Press.

Shaked, M. and Shanthikumar, J. (2007). *Stochastic Orders.* New York: Springer-Verlag.

Shapiro, A. (1990). “On concepts of directional differentiability”. *Journal of Optimization Theory and Applications* 66, pp. 477–487.

Shapiro, A. (1991). “Asymptotic analysis of stochastic programs”. *Annals of Operations Research* 30, pp. 169–186.

Sklar, A. (1959). “Fonctions de répartition à n dimensions et leurs marges”. *Publications de l’Institut de Statistique de L’Université de Paris* 8, pp. 229–231.

Snijders, T. (1981). “Rank tests for bivariate symmetry”. *Annals of Statistics* 9, pp. 1087–1095.

Vaart, A. van der and Wellner, J. (1996). *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

Yanagimoto, T. and Sibuya, M. (1972). “Stochastically larger component of a random vector”. *Annals of the Institute of Statistical Mathematics* 24, pp. 259–269.

Yanagimoto, T. and Sibuya, M. (1976). “Test of symmetry of a bivariate distribution”. *Sankhyā: The Indian Journal of Statistics, Series A* 38, pp. 105–115.

Appendix A

Supplementary materials for Chapter 2

A.1 Lemmas

We start with a general definition of a least star-shaped majorant functional. For any measurable set $E \subseteq [0,1]$ and fixed point $(c,d) \in \mathbb{R}^2$, where $c \geq \sup E$, we define the least star-shaped majorant of a function $h \in l([0,1])$ over E with kernel (c,d) by

$$\mathcal{M}_E^{(c,d)} h = \inf\{h^* \in l([0,1]) : h \leq h^* \text{ and } h^* \text{ is star-shaped over } E \text{ with kernel } (c,d)\}.$$

We say that $h^* \in l([0,1])$ is star-shaped over E with kernel (c,d) if $\{d - h^*(u)\}/(c-u)$ is nonincreasing over E.

When $c \in E$, the slope function $\{d - h^*(u)\}/(c-u)$ is not defined at $u = c$. To make this definition more precise, we say that $h^* \in l([0,1])$ is star-shaped over E with kernel (c,d) if

- when $c \notin E$, $\{d - h^*(u)\}/(c-u)$ is nonincreasing over E
- when $c \in E$, $\{d - h^*(u)\}/(c-u)$ is nonincreasing over $E \setminus \{c\}$ and $h^*(c) \geq d$.

This general definition includes $\mathcal{M} \equiv \mathcal{M}_{[0,1]}^{(1,1)}$ and $\mathcal{M}_{[a,b]}^{(1,0)}$, defined in Chapter 2, as special cases. Lemma A.1 shows how to calculate $\mathcal{M}_E^{(c,d)} h$ for any $h \in l([0,1])$.

Lemma A.1 (Calculation of $\mathcal{M}_E^{(c,d)} h$)**.** *For any function $h \in l([0,1])$, $\mathcal{M}_E^{(c,d)} h$ is* *star-shaped over E with kernel (c,d) and can be calculated by*

$$\mathcal{M}_E^{(c,d)}h(u) = \begin{cases} \max\{h(c), d\}, & \textit{if } c \in E \textit{ and } u = c \\ d - \gamma_E^h(u)(c-u), & \textit{if } u \in E \textit{ and } u \neq c \\ h(u), & \textit{if } u \notin E, \end{cases}$$

where $\gamma_E^h(u) = \inf_{v \le u, v \in E}\{d - h(v)\}/(c-v)$.

Proof of Lemma A.1. It is easy to see that $\mathcal{M}_E^{(c,d)}h(u) = h(u)$ when $u \notin E$. We now show that $\mathcal{M}_E^{(c,d)}h(c) = \max\{h(c), d\}$ when $c \in E$. For $h \in l([0,1])$, denote by

$$\mathcal{F}_h = \{h^* \in l([0,1])\colon\ h^* \ge h \text{ and } h^* \text{ is star-shaped over } E \text{ with kernel } (c,d)\}.$$

Because $\mathcal{M}_E^{(c,d)}h = \inf \mathcal{F}_h$, we have $\mathcal{M}_E^{(c,d)}h(c) \ge h(c)$. Furthermore, from the star-shaped definition above, we have $\inf_{h^* \in \mathcal{F}_h} h^*(c) \ge d$. Therefore, $\mathcal{M}_E^{(c,d)}h(c) \ge \max\{h(c), d\}$. Next, we show $\mathcal{M}_E^{(c,d)}h(c) \le \max\{h(c), d\}$. For each $h^* \in \mathcal{F}_h$, if $d \ge h(c)$, let h^{**} be a function satisfying $h^{**}(u) = h^*(u)$ for $u \neq c$ and $h^{**}(c) = d$. Then $h^{**} \ge h$ and $\{d - h^{**}(u)\}/(c-u) = \{d - h^*(u)\}/(c-u)$ is nonincreasing over $E \setminus \{c\}$ when $h^{**}(c) \ge d$. Thus, $h^{**} \in \mathcal{F}_h$. Similarly, if $d \le h(c)$, letting $h^{**}(u) = h^*(u)$ for $u \neq c$ and $h^{**}(c) = h(c) \ge d$ leads to $h^{**} \in \mathcal{F}_h$. Therefore, $\mathcal{M}_E^{(c,d)}h(c) \le \max\{h(c), d\}$ and hence $\mathcal{M}_E^{(c,d)}h(c) = \max\{h(c), d\}$ when $c \in E$.

We now show that $\mathcal{M}_E^{(c,d)}h$ is star-shaped over E with kernel (c,d). Let

$$S^*(u) = \frac{d - \mathcal{M}_E^{(c,d)}h(u)}{c-u}.$$

Because $\mathcal{M}_E^{(c,d)}h(c) = \max\{h(c), d\} \ge d$, it suffices to show that $S^*(u)$ is nonincreasing over $E \setminus \{c\}$. For $u \in E \setminus \{c\}$, there exists a sequence of star-shaped functions over E with kernel (c,d), say H_n, such that $H_n - \mathcal{M}_E^{(c,d)}h \ge 0$ and $\sup_{u \in E}|H_n(u) - \mathcal{M}_E^{(c,d)}h(u)| < 1/n$ for each n. Let

$$s_n^*(u) = \frac{d - H_n(u)}{c-u}$$

and note that $\lim_{n\to\infty} s_n^*(u) = S^*(u)$. Because each s_n^* is a nonincreasing function over $E \setminus \{c\}$, S^* is nonincreasing over $E \setminus \{c\}$ as well; i.e., $\mathcal{M}_E^{(c,d)}h(u)$ is star-shaped over E with kernel (c, d).

To finish the proof, we show that $\mathcal{M}_E^{(c,d)}h = \mathcal{M}_E^{\dagger}h$ for any $h \in l([0,1])$, where

$$\mathcal{M}_E^{\dagger}h(u) = \begin{cases} \max\{h(c), d\}, & \text{if } c \in E \text{ and } u = c \\ d - \gamma_E^h(u)(c-u), & \text{if } u \in E \text{ and } u \neq c \\ h(u), & \text{if } u \notin E. \end{cases}$$

Clearly, $\mathcal{M}_E^{\dagger}h$ is star-shaped over E with kernel (c, d) and $\mathcal{M}_E^{\dagger}h \geq h$. This implies $\mathcal{M}_E^{(c,d)}h \leq \mathcal{M}_E^{\dagger}h$. Furthermore, $\mathcal{M}_E^{(c,d)}h(u) = \mathcal{M}_E^{\dagger}h(u)$ when $c \in E$ and $u = c$ or when $u \notin E$. Thus, it suffices to show $\mathcal{M}_E^{(c,d)}h(u) \geq \mathcal{M}_E^{\dagger}h(u)$ when $u \in E \setminus \{c\}$. Because $S^*(u)$ is nonincreasing over $E \setminus \{c\}$, we can write $S^*(u) = \inf_{v \leq u, v \in E} S^*(v)$. Let

$$s(v) = \frac{d - h(v)}{c - v}, \quad \text{for } v \in E \setminus \{c\}.$$

Because $\mathcal{M}_E^{(c,d)}h(v) \geq h(v)$, it follows that $S^*(v) \leq s(v)$ for $v \in E \setminus \{c\}$. Therefore, $S^*(u) = \inf_{v \leq u, v \in E} S^*(v) \leq \inf_{v \leq u, v \in E} s(v) = \gamma_E^h(u)$ for $u \in E \setminus \{c\}$. Consequently, $\mathcal{M}_E^{(c,d)}h(u) = d - S^*(u)(c-u) \geq \mathcal{M}_E^{\dagger}h(u) = d - \gamma_E^h(u)(c-u)$ for $u \in E \setminus \{c\}$. $\square$

Lemma A.2. *For $h \in l([0,1])$, the operator $\mathcal{M}_E^{(c,d)}$ enjoys the following properties:*

(a) *For $\tau > 0$,*

$$\mathcal{M}_E^{(c,d)}(\tau h) = \tau \mathcal{M}_E^{(c,d/\tau)}h.$$

(b) *If ξ is a linear function, then*

$$\mathcal{M}_E^{(c,d)}(h + \xi) = \xi + \mathcal{M}_E^{(c,d-\xi(c))}h.$$

Proof of Lemma A.2. From Lemma A.1, it is easy to see that (a) and (b) hold when $u \notin E$ or when $c \in E$ and $u = c$. Also from Lemma A.1 and with $u \in E \setminus \{c\}$, we have

$$\begin{aligned}\mathcal{M}_E^{(c,d)}(\tau h)(u) &= d - \inf_{v\le u, v\in E}\left\{\frac{d-\tau h(v)}{c-v}\right\}(c-u)\\ &= \tau\left[\frac{d}{\tau} - \inf_{v\le u, v\in E}\left\{\frac{\frac{d}{\tau}-h(v)}{c-v}\right\}(c-u)\right] = \tau\mathcal{M}_E^{(c,d/\tau)}h(u),\end{aligned}$$

which establishes (a). To prove (b), note that if ξ is linear, we can write $\xi(u) = \xi(c) + s(u-c)$, where s is the slope of ξ and $|s| < \infty$. Therefore, when $u \in E \setminus \{c\}$,

$$\begin{aligned}&\mathcal{M}_E^{(c,d)}(h+\xi)(u)\\ =\ & d - \inf_{v\le u, v\in E}\left\{\frac{d-(h+\xi)(v)}{c-v}\right\}(c-u)\\ =\ & d - \inf_{v\le u, v\in E}\left\{\frac{d-\xi(c)-h(v)+\xi(c)-\xi(v)}{c-v}\right\}(c-u)\\ =\ & d-\xi(c)+\xi(c) - \inf_{v\le u, v\in E}\left[\frac{\{d-\xi(c)\}-h(v)}{c-v}+s\right](c-u)\\ =\ & \underbrace{\xi(c)+s(u-c)}_{=\ \xi(u)} + \underbrace{d-\xi(c)-\inf_{v\le u, v\in E}\left[\frac{\{d-\xi(c)\}-h(v)}{c-v}\right](c-u)}_{=\ \mathcal{M}_E^{(c,d-\xi(c))}h(u)},\end{aligned}$$

which establishes (b). □

Lemma A.3 (Lipschitz continuity)**.** *For any $h_1, h_2 \in l([0,1])$,*

$$\|\mathcal{M}_E^{(c,d)}h_1 - \mathcal{M}_E^{(c,d)}h_2\|_\infty \le \|h_1-h_2\|_\infty.$$

Proof of Lemma A.3. From Lemma A.1, $\mathcal{M}_E^{(c,d)}h_1(u) = h_1(u)$ and $\mathcal{M}_E^{(c,d)}h_2(u) = h_2(u)$ when $u \notin E$ and hence $\sup_{u\notin E}|\mathcal{M}_E^{(c,d)}h_1(u) - \mathcal{M}_E^{(c,d)}h_2(u)| = \sup_{u\notin E}|h_1(u) - h_2(u)| \le \|h_1-h_2\|_\infty$. Also, when $c \in E$ and $u = c$, $|\mathcal{M}_E^{(c,d)}h_1(u) - \mathcal{M}_E^{(c,d)}h_2(u)| = |\max\{h_1(c), d\} - \max\{h_2(c), d\}| \le |h_1(c)-h_2(c)| \le \|h_1-h_2\|_\infty$. Therefore, it suffices to show that for $u \in E\setminus\{c\}$,

$$|\mathcal{M}_E^{(c,d)}h_1(u) - \mathcal{M}_E^{(c,d)}h_2(u)| \le \sup_{u\in E\setminus\{c\}}|h_1(u)-h_2(u)|.$$

For fixed $u \in E\backslash\{c\}$, define $s_i(u) = \{d - h_i(u)\}/(c-u)$, for $i = 1, 2$, and $E_u = \{v : v \leq u, v \in E\backslash\{c\}\}$ so that $\mathcal{M}_E^{(c,d)} h_1(u) - \mathcal{M}_E^{(c,d)} h_2(u) = \{\inf_{v \in E_u} s_2(v) - \inf_{v \in E_u} s_1(v)\}(c - u)$. It suffices to show that for $u \in E\backslash\{c\}$,

$$\left| \left\{ \inf_{v \in E_u} s_1(v) - \inf_{v \in E_u} s_2(v) \right\} (c-u) \right| \leq \sup_{u \in E\backslash\{c\}} \left| \{s_1(u) - s_2(u)\}(c-u) \right|. \qquad \text{(A.1)}$$

If $\inf_{v \in E_u} s_1(v) = \inf_{v \in E_u} s_2(v)$, then (A.1) clearly holds. On the other hand, if $\inf_{v \in E_u} s_1(v) \neq \inf_{v \in E_u} s_2(v)$, it suffices to consider the $\inf_{v \in E_u} s_1(v) > \inf_{v \in E_u} s_2(v)$ case as the other case holds similarly. From the definition of infimum, there exists $v_n \in E_u$ such that $s_1(v_n) > \inf_{v \in E_u} s_1(v) - 1/n$ and $s_2(v_n) < \inf_{v \in E_u} s_1(v) + 1/n$ so that $s_1(v_n) - s_2(v_n) > \inf_{v \in E_u} s_1(v) - \inf_{v \in E_u} s_2(v) - 2/n$. As $0 \leq v_n \leq u < c \leq 1$ for all n, we have

$$\begin{aligned}
\{s_1(v_n) - s_2(v_n)\}(c - v_n) &> \left\{ \inf_{v \in E_u} s_1(v) - \inf_{v \in E_u} s_2(v) - \frac{2}{n} \right\} (c - v_n) \\
&> \left\{ \inf_{v \in E_u} s_1(v) - \inf_{v \in E_u} s_2(v) \right\} (c - v_n) - \frac{2}{n} \\
&> \left\{ \inf_{v \in E_u} s_1(v) - \inf_{v \in E_u} s_2(v) \right\} (c - u) - \frac{2}{n}.
\end{aligned}$$

Since $\inf_{v \in E_u} s_1(v) > \inf_{v \in E_u} s_2(v)$, we can make $\inf_{v \in E_u} s_1(v) - \inf_{v \in E_u} s_2(v) - 2/n > 0$ for n sufficiently large and also

$$\begin{aligned}
\left|\{s_1(v_n) - s_2(v_n)\}(c - v_n)\right| &> \left| \left\{ \inf_{v \in E_u} s_1(v) - \inf_{v \in E_u} s_2(v) \right\} (c - u) - \frac{2}{n} \right| \\
&\geq \left| \left\{ \inf_{v \in E_u} s_1(v) - \inf_{v \in E_u} s_2(v) \right\} (c - u) \right| - \frac{2}{n}.
\end{aligned}$$

Therefore, since $\left|\{s_1(v_n) - s_2(v_n)\}(c - v_n)\right| \leq \sup_{u \in E\backslash\{c\}} \left|\{s_1(u) - s_2(u)\}(c-u)\right|$ for all n, we have

$$\sup_{u \in E\backslash\{c\}} \left|\{s_1(u) - s_2(u)\}(c-u)\right| + \frac{2}{n} \geq \left| \left\{ \inf_{v \in E_u} s_1(v) - \inf_{v \in E_u} s_2(v) \right\} (c - u) \right|.$$

Taking limits as $n \to \infty$ completes the proof. $\square$

We now investigate the Hadamard directional differentiability properties of our operators $\mathcal{M} \equiv \mathcal{M}_{[0,1]}^{(1,1)}$ and $\mathcal{D} = \mathcal{M} - \mathcal{I}$, both defined in Chapter 2. The following definition is adapted from Shapiro (1990) and Shapiro (1991).

Definition A.1. Let $\mathcal{T}$ denote a map from $\mathcal{X}$ to $\mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are normed spaces with norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Y}}$, respectively. We say $\mathcal{T}$ is Hadamard directionally differentiable at $R \in \mathcal{X}$ tangentially to $\mathcal{X}_0 \subset \mathcal{X}$ if there exists a map $d\mathcal{T}_R : \mathcal{X}_0 \to \mathcal{Y}$ such that for all $h \in \mathcal{X}_0$,

$$\lim_{n\to\infty} \left\| t_n^{-1}\{\mathcal{T}(R + t_n h_n) - \mathcal{T}R\} - d\mathcal{T}_R h \right\|_{\mathcal{Y}} = 0$$

whenever $\|h_n - h\|_{\mathcal{X}} \to 0$ and $t_n \downarrow 0$ as $n \to \infty$.

Remark A.1. *In Chapter 2, we choose $\mathcal{X} = \mathcal{Y} = l([0,1])$ equipped with the sup-norm $\|\cdot\|_\infty$ and $\mathcal{X}_0 = C([0,1])$, the collection of all real continuous functions with domain $[0,1]$. Beare and Moon (2015), in their investigation of likelihood ratio ordering, take $\mathcal{T}$ to be the least concave majorant operator and show that $\mathcal{T}$ is Hadamard directionally differentiable at R (R concave) tangentially to $\mathcal{X}_0$. Unfortunately, in our investigation of uniform stochastic ordering, the operators $\mathcal{M}$ and $\mathcal{D}$ are* not *Hadamard directionally differentiable at all R (R star-shaped). To provide more details, we first state and prove the next lemma which helps us to characterize $d\mathcal{M}_R$ and $d\mathcal{D}_R$ from a pointwise convergence point of view.*

Lemma A.4. *Suppose $R \in \Theta_0$, possibly non-strictly star-shaped over closed, pairwise disjoint intervals of the form $[a_k, b_k]$, $0 \le a_k < b_k \le 1$, as described in Section 2.3.1. Suppose $h \in C([0,1])$, $h_n \in l([0,1])$, and $u \in [0,1]$. When $\|h_n - h\|_\infty \to 0$ and $t_n \downarrow 0$ as $n \to \infty$, then*

$$\begin{aligned} \lim_{n\to\infty} t_n^{-1}\left\{\mathcal{M}(R + t_n h_n) - \mathcal{M}R\right\}(u) &= d\mathcal{M}_R h(u) \\ \lim_{n\to\infty} t_n^{-1}\left\{\mathcal{D}(R + t_n h_n) - \mathcal{D}R\right\}(u) &= d\mathcal{D}_R h(u), \end{aligned}$$

where

$$d\mathcal{M}_R h(u) = \begin{cases} \max\{h(1), 0\}, & \text{if } u = 1 \\ \mathcal{M}^{(1,0)}_{[a_k,b_k]} h(u), & \text{if } \exists k \text{ such that } a_k \le u \le b_k \\ h(u), & \text{otherwise} \end{cases}$$

and $d\mathcal{D}_R h = d\mathcal{M}_R h - h$.

Proof of Lemma A.4. For $p \in [1, \infty]$, provided that $d\mathcal{M}_R h$ exists, we have

$$\begin{aligned} d\mathcal{D}_R h &= \lim_{n\to\infty} t_n^{-1}\{\mathcal{D}(R + t_n h_n) - \mathcal{D}R\} \\ &= \lim_{n\to\infty} t_n^{-1}\{\mathcal{M}(R + t_n h_n) - \mathcal{M}R - t_n h_n\} \\ &= \lim_{n\to\infty} t_n^{-1}\{\mathcal{M}(R + t_n h_n) - \mathcal{M}R\} - \lim_{n\to\infty} h_n = d\mathcal{M}_R h - h. \end{aligned}$$

We now prove the existence of $d\mathcal{M}_R h$. From Lemma A.3, we have

$$\begin{aligned} &\left\| t_n^{-1}\left\{\mathcal{M}(R + t_n h_n) - \mathcal{M}R\right\} - t_n^{-1}\left\{\mathcal{M}(R + t_n h) - \mathcal{M}R\right\}\right\|_\infty \\ &= \left\| t_n^{-1}\mathcal{M}(R + t_n h_n) - t_n^{-1}\mathcal{M}(R + t_n h)\right\|_\infty \le t_n^{-1}\|t_n h_n - t_n h\|_\infty = \|h_n - h\|_\infty. \quad \text{(A.2)} \end{aligned}$$

Because $\|h_n - h\|_\infty \to 0$ as $n \to \infty$, the sup-norm difference between $t_n^{-1}\{\mathcal{M}(R + t_n h_n) - \mathcal{M}R\}$ and $t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\}$ becomes negligible as $n \to \infty$. Thus, we need only focus on $t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\}$ and hence proceed to show that $t_n^{-1}\{\mathcal{M}(R + t_n h)(u) - \mathcal{M}R(u)\}$ converges pointwise to $d\mathcal{M}_R h(u)$ for all $u \in [0, 1]$. When $u = 1$, since $t_n > 0$, we have

$$t_n^{-1}\{\mathcal{M}(R + t_n h)(1) - \mathcal{M}R(1)\} = t_n^{-1}(\max\{1 + t_n h(1), 1\} - 1) = \max\{h(1), 0\}.$$

For each $u \in [0, 1)$, let ξ_u denote the secant line that passes through $(1, 1)$ and $(u, \mathcal{M}R(u))$. Therefore, $\mathcal{M}R(u) = \xi_u(u)$ and thus

$$t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\}(u) = t_n^{-1}\{\mathcal{M}(R + t_n h) - \xi_u\}(u).$$

From Lemma A.2, we can rewrite the last equation as

$$t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\}(u) = \mathcal{M}^{(1,0)}_{[0,1]}\{t_n^{-1}(R - \xi_u) + h\}(u) = \mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u), \quad \text{(A.3)}$$

where $h^R_{n,u} = t_n^{-1}(R - \xi_u) + h$. In addition, from Lemma A.1 we know that for any $u \in [a_k, b_k]$, we can find $a_u \le u$ such that $R(v) = 1 - R'(a_u)(1 - v)$ for all $v \in [a_u, u]$; specifically, we can take $a_u = a_k$. When $u \notin [a_k, b_k]$ for any k, we set $a_u = u$. Thus,

$$\mathcal{M}^{(1,0)}_{[a_u,u]} h(u) = \begin{cases} \mathcal{M}^{(1,0)}_{[a_k,u]} h(u) = \mathcal{M}^{(1,0)}_{[a_k,b_k]} h(u), & \text{if } u \in [a_k, b_k] \\ \mathcal{M}^{(1,0)}_{[u,u]} h(u) = h(u), & \text{otherwise.} \end{cases}$$

Therefore, it suffices to show $\lim_{n\to\infty} \mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u) = \mathcal{M}^{(1,0)}_{[a_u,u]} h(u)$ for each $u \in [0,1)$. Suppose $\delta > 0$. Applying the triangle inequality yields

$$\begin{aligned}
&\left|\mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u) - \mathcal{M}^{(1,0)}_{[a_u,u]} h(u)\right| \\
&\quad \le \left|\mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u) - \mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]} h^R_{n,u}(u)\right| + \left|\mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]} h^R_{n,u}(u) - \mathcal{M}^{(1,0)}_{[a_u,u]} h(u)\right|,
\end{aligned} \tag{A.4}$$

where $a \vee b = \max\{a,b\}$. Concentrating on the first term on the right-hand side of (A.4), we first show that

$$\lim_{n\to\infty} \left|\mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]} h^R_{n,u}(u) - \mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u)\right| = 0. \tag{A.5}$$

Remark A.2. *For large n, Equation* (A.5) *says that $\mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u)$ can be approximated by $\mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]} h^R_{n,u}(u)$. That is, the region $[0 \vee (a_u - \delta), u]$ provides enough information to determine $\mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u)$ when n is sufficiently large regardless of the value of $h^R_{n,u}(v)$ for $v > u$. Before moving forward, we examine the relationship between $h^R_{n,u}$ and h. From Lemma A.1, there exists $a_u \in [0,u]$ such that, for all $v \in [a_u,u]$, $R(v) = 1 - R'(a_u)(1-v)$ which coincides with the secant line $\xi_u(v)$; i.e., when $v \in [a_u,u]$, we have $R(v) = \xi_u(v)$ and thus $h^R_{n,u}(v) = h(v)$. When $v \in [0,a_u)$, we have $h^R_{n,u}(v) < h(v)$ because R is below the secant line ξ_u over $[0,a_u)$.*

We now proceed to show that (A.5) holds. If $\delta \ge a_u$, then (A.5) holds trivially because

$$\mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]} h^R_{n,u}(u) = \mathcal{M}^{(1,0)}_{[0,u]} h^R_{n,u}(u) = -\inf_{v\le u}\left\{\frac{-h^R_{n,u}(v)}{1-v}\right\}(1-u) = \mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u).$$

For $0 < \delta < a_u$, consider the slopes of the linear segment from $(v, h^R_{n,u}(v))$ to $(1,0)$ for all $v \in [0,u]$ with $u \in (0,1)$ and $a_u \in (\delta,u]$. Note that for each fixed $v \in [0,a_u)$, $R(v) - \xi_u(v) < 0$. Also, because t_n^{-1} diverges, so does $\{-h^R_{n,u}(v)\}/(1-v) = -\{t_n^{-1}(R-\xi_u) + h\}(v)/(1-v)$ as $n \to \infty$. Thus, there exists an $N_1(u,v,\delta) \ge 1$ such that for all

$n \geq N_1(u, v, \delta)$,

$$\frac{-h_{n,u}^R(v)}{1-v} > \frac{-h(u)}{1-u} = \frac{-h_{n,u}^R(u)}{1-u}. \tag{A.6}$$

We can also conclude that there exists an $N_2(u, \delta) \geq 1$ for all $v \in [0, a_u - \delta]$ such that

$$\inf_{v \in [0, a_u - \delta]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\} \geq \frac{-h_{n,u}^R(u)}{1-u} \tag{A.7}$$

for all $n \geq N_2(u, \delta)$. To see why (A.7) holds, define

$$E_n = \left\{ v: \ v < a_u, \ \frac{-h_{n,u}^R(v)}{1-v} > \frac{-h_{n,u}^R(u)}{1-u} \right\}.$$

From (A.6), we can see that E_n is an increasing sequence of sets in n for fixed $u \in [0, 1)$ and $\delta \in (0, a_u)$. Since R and h are continuous, $h_{n,u}^R(v)$ is continuous over $[0, a_u]$. Therefore, $h_{n,u}^R(v)/(1-v)$ is uniformly continuous over $[0, a_u]$ and thus E_n is open relative to $[0, 1]$. That is, $\{E_n\}_{n=1}^{\infty}$ forms an open covering, relative to $[0, 1]$, of the compact set $[0, a_u - \delta]$. By the Heine-Borel theorem, there exists a finite collection of the E_n sets that cover $[0, a_u - \delta]$. In other words, there exists a finite $N_2(u, \delta) \geq 1$ such that (A.7) holds for all $n \geq N_2(u, \delta)$. Note that

$$\inf_{v \in [0, u]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\} = \min \left\{ \inf_{v \in [0, a_u - \delta]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\}, \inf_{v \in (a_u - \delta, u]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\} \right\}.$$

From (A.7), we can say that when $n \geq N_2(u, \delta)$,

$$\inf_{v \in [0, a_u - \delta]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\} \geq \frac{-h_{n,u}^R(u)}{1-u} \geq \inf_{v \in [0, u]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\}.$$

Therefore, for n sufficiently large,

$$\inf_{v \in (a_u - \delta, u]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\} = \inf_{v \in [0, u]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\},$$

and, by Lemma A.1, the left-hand side of (A.5) equals

$$\begin{aligned} &\lim_{n \to \infty} \left| \inf_{v \in (a_u - \delta, u]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\} (1-u) - \inf_{v \in [0, u]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\} (1-u) \right| \\ &\qquad = \lim_{n \to \infty} \left| \inf_{v \in (a_u - \delta, u]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\} - \inf_{v \in [0, u]} \left\{ \frac{-h_{n,u}^R(v)}{1-v} \right\} \right| (1-u) = 0. \end{aligned}$$

This shows that (A.5) holds.

We now consider the second term on the right-hand side of (A.4). For $\delta > 0$, it is easy to see from Remark A.2 that

$$\mathcal{M}^{(1,0)}_{[a_u,u]}h(u) = \mathcal{M}^{(1,0)}_{[a_u,u]}h^R_{n,u}(u) \leq \mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]}h^R_{n,u}(u) \leq \mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]}h(u)$$

for all n and thus

$$|\mathcal{M}^{(1,0)}_{[a_u,u]}h(u) - \mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]}h^R_{n,u}(u)| \leq \mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]}h(u) - \mathcal{M}^{(1,0)}_{[a_u,u]}h(u).$$

Therefore, to finish the proof of Lemma A.4, it suffices to show that

$$\lim_{\delta\to 0}\left\{\mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]}h(u) - \mathcal{M}^{(1,0)}_{[a_u,u]}h(u)\right\} = 0.$$

If $a_u = 0$, then this holds trivially. If $a_u > 0$, take $\delta \in (0, a_u)$. Define

$$H_u(v) = \left\{\frac{-h(v)}{1-v}\right\}(1-u), \quad \text{for } 0 \leq v \leq u < 1,$$

so that the difference $\mathcal{M}^{(1,0)}_{[0\vee(a_u-\delta),u]}h(u) - \mathcal{M}^{(1,0)}_{[a_u,u]}h(u)$ can be written as the difference of two supremums $\sup_{v\in[a_u-\delta,u]} H_u(v) - \sup_{v\in[a_u,u]} H_u(v)$. Since $H_u(v)$ is uniformly continuous over $[0, u]$, then for any $\epsilon > 0$, there exists a $\delta^*(\epsilon, u) > 0$, such that $|H_u(v) - H_u(a_u)| < \epsilon$ for all $v \in [a_u - \delta, a_u)$ whenever $\delta < \delta^*(\epsilon, u)$. Thus, $\sup_{v\in[a_u-\delta,a_u)} H_u(v) \leq H_u(a_u) + \epsilon$ when $\delta < \delta^*(\epsilon, u)$ and also

$$\begin{aligned}
&\sup_{v\in[a_u-\delta,u]} H_u(v) - \sup_{v\in[a_u,u]} H_u(v) \\
&\qquad = \max\left\{\sup_{v\in[a_u-\delta,a_u)} H_u(v), \sup_{v\in[a_u,u]} H_u(v)\right\} - \sup_{v\in[a_u,u]} H_u(v) \\
&\qquad \leq \max\left\{H_u(a_u) + \epsilon, \sup_{v\in[a_u,u]} H_u(v)\right\} - \sup_{v\in[a_u,u]} H_u(v) \\
&\qquad \leq \sup_{v\in[a_u,u]} H_u(v) + \epsilon - \sup_{v\in[a_u,u]} H_u(v) = \epsilon
\end{aligned}$$

when $\delta < \delta^*(\epsilon, u)$. This completes the proof. □

Remark A.3. *An anonymous reader has suggested that developing pointwise confidence intervals for $R(u)$ under a uniform stochastic ordering constraint may be an* *interesting pursuit. Lemma A.4 could serve as a starting point towards accomplishing this. To see how, note that when $R \in \Theta_0$; i.e., uniform stochastic ordering holds, we can use Lemma A.4 and Skorohod's representation theorem to write*

$$c_{mn}\{\mathcal{M}R_{mn}(u) - R(u)\} \xrightarrow{d} d\mathcal{M}_R T_R^\lambda(u), \quad \textit{for } u \in [0,1],$$

as $\min\{m,n\} \to \infty$ *and* $n/(m+n) \to \lambda \in (0,1)$, *where the process* T_R^λ *is defined in the proof of Theorem 2.1. The challenge going forward would be how exactly to approximate the distribution of* $d\mathcal{M}_R T_R^\lambda(u)$. *This could be rather formidable because* $d\mathcal{M}_R T_R^\lambda(u)$ *depends on* λ, R *itself, and the derivative* R'. *It also inherently depends on the non-strictly star-shaped regions* $[a_k, b_k]$, *for* $k = 1, 2, \ldots$. *We believe one could approximate an upper bound of* $d\mathcal{M}_R T_R^\lambda(u)$ *that avoids having to locate these regions, but the resulting confidence interval for* $R(u)$ *could be very conservative.*

Remark A.4. *To show that both* $\mathcal{M}$ *and* $\mathcal{D}$ *are Hadamard directionally differentiable at* R *tangentially to* $C([0,1])$, *Lemma A.4 is not strong enough. Lemma A.4 describes pointwise convergence but uniform convergence is needed; i.e.,*

$$\begin{aligned} \lim_{n\to\infty} \|t_n^{-1}\{\mathcal{M}(R + t_n h_n) - \mathcal{M}R\} - d\mathcal{M}_R h\|_\infty &= 0 \\ \lim_{n\to\infty} \|t_n^{-1}\{\mathcal{D}(R + t_n h_n) - \mathcal{D}R\} - d\mathcal{D}_R h\|_\infty &= 0. \end{aligned}$$

However, uniform convergence does not hold for all $R \in \Theta_0$. *It does hold when* $R \in \Theta_0$ *satisfies certain conditions, as we now describe.*

Lemma A.5 (Hadamard directional differentiability)**.** *If* $R \in \Theta_0$ *has at most one non-strictly star-shaped region* $[a_1, 1]$ *where* $0 \le a_1 < 1$, *then* $\mathcal{M}$ *and* $\mathcal{D}$ *are both Hadamard directionally differentiable at* R *tangentially to* $C([0,1])$ *with derivatives* $d\mathcal{M}_R$ *and* $d\mathcal{D}_R$, *respectively. Otherwise,* $\mathcal{M}$ *and* $\mathcal{D}$ *are not Hadamard directionally differentiable.*

Proof of Lemma A.5. We use the notation from the proof of Lemma A.4. When $R \in \Theta_0$ has at most one non-strictly star-shaped region of the form $[a_1, 1]$, $d\mathcal{M}_R h$ is a continuous function on $[0,1]$. Furthermore, for any $h \in C([0,1])$, the sequence of continuous functions $t_n^{-1}\{\mathcal{M}(R+t_nh)-\mathcal{M}R\}$ converges pointwise to $d\mathcal{M}_Rh$, and, in Lemma A.6, we show that convergence is monotone. Applying Dini's theorem yields $\lim_{n\to\infty}\|t_n^{-1}\{\mathcal{M}(R+t_nh)-\mathcal{M}R\}-d\mathcal{M}_Rh\|_\infty=0$. By (A.2), we have

$$\lim_{n\to\infty}\|t_n^{-1}\{\mathcal{M}(R+t_nh_n)-\mathcal{M}R\}-d\mathcal{M}_Rh\|_\infty=0.$$

Similarly, $\lim_{n\to\infty}\|t_n^{-1}\{\mathcal{D}(R+t_nh_n)-\mathcal{D}R\}-d\mathcal{D}_Rh\|_\infty=0$. When $R\in\Theta_0$ has at least one non-strictly star-shaped region $[a_k,b_k]$, where $b_k<1$, then $\mathcal{M}$ is not Hadamard directionally differentiable at R tangentially to $C([0,1])$; i.e., there exists an $h\in C([0,1])$ such that $\lim_{n\to\infty}\|t_n^{-1}\{\mathcal{M}(R+t_nh)-\mathcal{M}R\}-d\mathcal{M}_Rh\|_\infty\neq 0$ (and similarly for $\mathcal{D}$). We now provide an example to show this. Suppose

$$h(u)=\begin{cases}1-\dfrac{u}{b_k}, & \text{if } u\in[0,b_k)\\[1ex] 0, & \text{if } u\in[b_k,1].\end{cases}$$

From the definition of $d\mathcal{M}_Rh$, we know that $\mathcal{M}^{(1,0)}_{[a_k,b_k]}h(b_k)-h(b_k)=\Delta$, where $\Delta=(b_k-a_k)(1-b_k)/\{b_k(1-a_k)\}$ and that there exists a $\delta>0$, such that $d\mathcal{M}_Rh(u)=h(u)$ when $u\in(b_k,b_k+\delta)$. Because both $\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,u}(u)$ and $h(u)$ are continuous functions of u, for any $\epsilon>0$, we can find a $\delta^*(\epsilon,b_k)>0$ such $|\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,w}(w)-\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,b_k}(b_k)|<\epsilon/3$ and $|h(b_k)-h(w)|<\epsilon/3$ whenever $0<w-b_k<\delta^*(\epsilon,b_k)$. Since $\lim_{n\to\infty}\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,u}(u)=\mathcal{M}^{(1,0)}_{[a_k,u]}h(u)$ for each $u\in[0,1)$, we can find an $N(\epsilon,b_k)\geq 1$ such that $|\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,b_k}(b_k)-\mathcal{M}^{(1,0)}_{[a_k,b_k]}h(b_k)|<\epsilon/3$ when $n\geq N(\epsilon,b_k)$. Therefore, for any w satisfying $b_k<w<b_k+\min\{\delta,\delta^*(\epsilon,b_k)\}$,

$$\begin{aligned}\left\|t_n^{-1}\{\mathcal{M}(R+t_nh)-\mathcal{M}R\}-d\mathcal{M}_Rh\right\|_\infty &\geq \sup_{u\in[0,1)}\left|\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,u}(u)-d\mathcal{M}_Rh(u)\right|\\ &\geq \left|\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,w}(w)-h(w)\right|\end{aligned}$$

which is greater than or equal to

$$\begin{aligned}&\left|\mathcal{M}^{(1,0)}_{[a_k,b_k]}h(b_k)-h(b_k)\right|-\left|\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,w}(w)-\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,b_k}(b_k)\right|\\ &\qquad-\left|\mathcal{M}^{(1,0)}_{[0,1]}h^R_{n,b_k}(b_k)-\mathcal{M}^{(1,0)}_{[a_k,b_k]}h(b_k)\right|-|h(b_k)-h(w)|\end{aligned}$$

which is greater than or equal to $|\Delta| - \epsilon$ for all $n \geq N(\epsilon, b_k)$. Therefore, for this choice of h, we have shown $\lim_{n\to\infty} \|t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\} - d\mathcal{M}_R h\|_\infty > 0$. A similar result involving $\mathcal{D}$ can also be shown. $\square$

Lemma A.6. *Suppose $R \in \Theta_0$. Under the same conditions stated in Lemma A.4, it follows that for all $p \in [1, \infty]$,*

$$\begin{aligned}\lim_{n\to\infty} \left\| t_n^{-1} \{\mathcal{M}(R + t_n h_n) - \mathcal{M}R\} \right\|_p &= \|d\mathcal{M}_R h\|_p \\ \lim_{n\to\infty} \left\| t_n^{-1} \{\mathcal{D}(R + t_n h_n) - \mathcal{D}R\} \right\|_p &= \|d\mathcal{D}_R h\|_p .\end{aligned}$$

Remark A.5. *When $\mathcal{M}$ and $\mathcal{D}$ are not Hadamard directionally differentiable, Lemma A.6 plays a crucial role in the proofs of Theorems 2.1 and 2.2. In Lemma A.4, we have shown that $t_n^{-1}\{\mathcal{M}(R+t_n h_n)-\mathcal{M}R\}$ and $t_n^{-1}\{\mathcal{D}(R+t_n h_n)-\mathcal{D}R\}$ converge pointwise to $d\mathcal{M}_R h$ and $d\mathcal{D}_R h$, respectively. However, pointwise convergence does not guarantee that the L^p norms converge to the corresponding L^p norms of the limits. To prove Lemma A.6, we consider the $p \in [1, \infty)$ and $p = \infty$ cases separately. When $p < \infty$, the proof is straightforward as we can appeal to the monotone convergence theorem. However, when $p = \infty$, no such tools can be used. A large amount of effort in the proof of Lemma A.6 is expended on the $p = \infty$ case.*

Proof of Lemma A.6. We use the same notation from the proof of Lemma A.4. From (A.2), we have $\|t_n^{-1}\{\mathcal{M}(R+t_n h_n)-\mathcal{M}R\} - t_n^{-1}\{\mathcal{M}(R+t_n h)-\mathcal{M}R\}\|_p \leq \|h_n - h\|_\infty$ and $\|t_n^{-1}\{\mathcal{D}(R + t_n h_n) - \mathcal{D}R\} - t_n^{-1}\{\mathcal{D}(R + t_n h) - \mathcal{D}R\}\|_p \leq 2\|h_n - h\|_\infty$. Therefore, it suffices to show that for all $p \in [1, \infty]$,

$$\lim_{n\to\infty} \left\| t_n^{-1} \{\mathcal{M}(R + t_n h) - \mathcal{M}R\} \right\|_p = \|d\mathcal{M}_R h\|_p \tag{A.8}$$

and

$$\lim_{n\to\infty} \left\| t_n^{-1} \{\mathcal{D}(R + t_n h) - \mathcal{D}R\} \right\|_p = \|d\mathcal{D}_R h\|_p . \tag{A.9}$$

Let $f_n(u) = t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\}(u)$ and $f(u) = d\mathcal{M}_R h(u)$ for $u \in [0, 1]$. From Lemma A.4, we know that f_n converges pointwise to f, $f_n(1) = f(1) = \max\{h(1), 0\}$, and $f_n(u) = \mathcal{M}_{[0,1]}^{(1,0)} h_{n,u}^R(u)$ for $u \in [0,1)$. Suppose $u \in [0,1)$. Suppose n_1 and n_2 are positive integers satisfying $n_1 < n_2$. From the definition of $\mathcal{M}_{[0,1]}^{(1,0)} h_{n,u}^R(u)$, we know $R(v) - \xi_u(v) \leq 0$ for $v \leq u$. Since $t_n \downarrow 0$, we have $-t_{n_1}^{-1}\{R(v) - \xi_u(v)\} \leq -t_{n_2}^{-1}\{R(v) - \xi_u(v)\}$ and

$$\begin{aligned}
\mathcal{M}_{[0,1]}^{(1,0)} h_{n_1,u}^R(u) &= -\inf_{v \leq u} \left\{ \frac{-h_{n_1,u}^R(v)}{1-v} \right\} (1-u) \\
&= -\inf_{v \leq u} \left[\frac{-t_{n_1}^{-1}\{R(v) - \xi_u(v)\} - h(v)}{1-v} \right] (1-u) \\
&\geq -\inf_{v \leq u} \left[\frac{-t_{n_2}^{-1}\{R(v) - \xi_u(v)\} - h(v)}{1-v} \right] (1-u) \\
&= -\inf_{v \leq u} \left\{ \frac{-h_{n_2,u}^R(v)}{1-v} \right\} (1-u) = \mathcal{M}_{[0,1]}^{(1,0)} h_{n_2,u}^R(u).
\end{aligned}$$

Therefore, for each $u \in [0,1]$, it follows that $\{f_n(u)\}_{n=1}^{\infty}$ is a nonincreasing sequence that converges to $f(u)$.

When $1 \leq p < \infty$, a direct application of Lebesgue's monotone convergence theorem implies that $\lim_{n\to\infty} \|f_n - f\|_p = 0$, that is, $\lim_{n\to\infty} \|t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\} - d\mathcal{M}_R h\|_p = 0$. From Lemma A.4, we know that $d\mathcal{D} = d\mathcal{M} - \mathcal{I}$. It is easy to see that $\|t_n^{-1}\{\mathcal{D}(R + t_n h) - \mathcal{D}R\} - d\mathcal{D}h\|_p = \|t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\} - d\mathcal{M}h\|_p$, for $1 \leq p < \infty$. Therefore, $\lim_{n\to\infty} \|t_n^{-1}\{\mathcal{D}(R + t_n h) - \mathcal{D}R\} - d\mathcal{D}_R h\|_p = 0$. This establishes (A.8) and (A.9), respectively, when p is finite.

When $p = \infty$, we first prove that (A.8) holds, that is,

$$\lim_{n\to\infty} \left\| t_n^{-1} \left\{ \mathcal{M}(R + t_n h) - \mathcal{M}R \right\} \right\|_\infty = \|d\mathcal{M}_R h\|_\infty$$

or, equivalently, $\lim_{n\to\infty} \|f_n\|_\infty = \|f\|_\infty$. We first show $\liminf_{n\to\infty} \|f_n\|_\infty \geq \|f\|_\infty$. Given $\epsilon > 0$, there exists an x_ℓ such that $|f(x_\ell)| > \|f\|_\infty - \epsilon/2$. Since $|f_n|$ converges pointwise to $|f|$, there exists an $N(\epsilon, x_\ell) \geq 1$ such that $-\epsilon/2 < |f_n(x_\ell)| - |f(x_\ell)| < \epsilon/2$ for all $n \geq N(\epsilon, x_\ell)$ and therefore

$$\|f_n\|_\infty \geq |f_n(x_\ell)| > |f(x_\ell)| - \frac{\epsilon}{2} > \|f\|_\infty - \epsilon.$$

Since $\epsilon > 0$ is arbitrary, $\liminf_{n\to\infty} \|f_n\|_\infty \geq \|f\|_\infty$. We now show $\limsup_{n\to\infty} \|f_n\|_\infty \leq \|f\|_\infty$. Suppose the non-strictly star-shaped region of R takes the form $\cup_{k=1}^{K} [a_k, b_k]$, where $K < \infty$ and $0 = a_0 = b_0 \leq a_1 < b_1 < \cdots < a_K \leq b_K \leq a_{K+1} = 1$. For $h \in C([0,1])$, we construct the auxiliary function h^*, defined by

$$h^*(u) = \begin{cases} \max\left\{\sup_{\ell<k}\left\{\mathcal{M}^{(1,0)}_{[a_\ell,b_\ell]}h(b_\ell)\left(\dfrac{1-u}{1-b_\ell}\right)\right\},\ \mathcal{M}^{(1,0)}_{[a_k,b_k]}h(u)\right\}, & \exists k, u \in [a_k, b_k] \\ \max\left\{\sup_{\ell\leq k}\left\{\mathcal{M}^{(1,0)}_{[a_\ell,b_\ell]}h(b_\ell)\left(\dfrac{1-u}{1-b_\ell}\right)\right\},\ h(u)\right\}, & \exists k, u \in (b_k, a_{k+1}], \end{cases}$$

We force $\sup_{l<0}\{\mathcal{M}^{(1,0)}_{[a_\ell,b_\ell]}h(b_\ell)(1-u_\ell)/(1-b_\ell)\} = -\infty$; i.e., $h^*(0) = h(0)$.

Remark A.6. *Note that this choice of h^* is critical. In particular, it satisfies the following three properties:*

C1. $h^* \in C([0,1])$

C2. $d\mathcal{M}_R h^* = h^*$

C3. $h^* \geq h$.

A slightly modified version of h^* is available when R has infinitely many non-strictly star-shaped regions. This modified version also satisfies the three properties above.

Property C1 allows us to use results from the proof of Lemma A.4. Since h^* is also a member of $C([0,1])$ and defining $f_n^*(u) = t_n^{-1}\{\mathcal{M}(R + t_n h^*) - \mathcal{M}R\}(u)$, for $u \in [0,1]$, it follows that f_n^* converges pointwise to $d\mathcal{M}_R h^*$ in $[0,1]$. Furthermore, because both $\mathcal{M}(R + t_n h^*)$ and $\mathcal{M}R$ are continuous, so is f_n^*. Property C2 says $d\mathcal{M}_R h^*(u) = h^*(u)$ is also continuous. Thus, we can apply Dini's theorem to obtain $\lim_{n\to\infty}\|f_n^*\|_\infty = \|h^*\|_\infty$. Property C3 implies $f \leq f_n \leq f_n^*$. It follows that $\limsup_{n\to\infty}\|f_n\|_\infty \leq \max\{\lim_{n\to\infty}\|f_n^*\|_\infty, \|f\|_\infty\} = \max\{\|h^*\|_\infty, \|f\|_\infty\}$. Therefore, to show $\limsup_{n\to\infty}\|f_n\|_\infty \leq \|f\|_\infty$, it suffices to show $\|f\|_\infty = \|d\mathcal{M}_R h\|_\infty \geq \|h^*\|_\infty$. If u is in one of the non-strictly star-shaped regions in $\cup_k[a_k, b_k]$, it follows that

$$\begin{aligned} |h^*(u)| &\leq \max\left\{\sup_{\ell<k}\left\{\left|\mathcal{M}^{(1,0)}_{[a_\ell,b_\ell]}h(b_\ell)\right|\left(\frac{1-u}{1-b_\ell}\right)\right\}, \left|\mathcal{M}^{(1,0)}_{[a_k,b_k]}h(u)\right|\right\} \\ &\leq \max\left\{\sup_{\ell<k}\left\{\left|\mathcal{M}^{(1,0)}_{[a_\ell,b_\ell]}h(b_\ell)\right|\right\}, \left|\mathcal{M}^{(1,0)}_{[a_k,b_k]}h(u)\right|\right\} \leq \sup_k \sup_{u\in[a_k,b_k]}\left|\mathcal{M}^{(1,0)}_{[a_k,b_k]}h(u)\right|. \end{aligned}$$

If u is in the strictly star-shaped region $(\cup_k[a_k,b_k])^c$, then

$$\begin{aligned}
|h^*(u)| &\leq \max\left\{\sup_{\ell<k}\left\{\left|\mathcal{M}^{(1,0)}_{[a_\ell,b_\ell]}h(b_\ell)\right|\left(\frac{1-u}{1-b_\ell}\right)\right\}, |h(u)|\right\} \\
&\leq \max\left\{\sup_{\ell<k}\left\{\left|\mathcal{M}^{(1,0)}_{[a_\ell,b_\ell]}h(b_\ell)\right|\right\}, |h(u)|\right\} \\
&\leq \max\left\{\sup_k \sup_{u\in[a_k,b_k]}\left|\mathcal{M}^{(1,0)}_{[a_k,b_k]}h(u)\right|, \sup_{u\notin\cup_k[a_k,b_k]}|h(u)|\right\}.
\end{aligned}$$

Combining

$$\|d\mathcal{M}_R h\|_\infty = \sup_{u\in[0,1]}|d\mathcal{M}_R h(u)| = \max\left\{\sup_k \sup_{u\in[a_k,b_k]}\left|\mathcal{M}^{(1,0)}_{[a_k,b_k]}h(u)\right|, \sup_{u\notin\cup_k[a_k,b_k]}|h(u)|\right\}$$

with the last two results shows that $\|h^*\|_\infty \leq \|d\mathcal{M}_R h\|_\infty$. We have shown that (A.8) holds in the $p=\infty$ case.

We now show (A.9) holds in the $p=\infty$ case. For $u\in[0,1]$, redefine $f_n(u) = t_n^{-1}\{\mathcal{M}(R+t_nh)-\mathcal{M}R\}(u)-h(u)$ and $f(u)=d\mathcal{D}_R h(u)$. Showing $\liminf_{n\to\infty}\|f_n\|_\infty \geq \|f\|_\infty$ is done the same way as before. To show $\limsup_{n\to\infty}\|f_n\|_\infty \leq \|f\|_\infty$, consider $h^* = h + h^D$ in place of h^* as previously defined, where

$$h^D(u) = \begin{cases} \mathcal{D}^{(1,0)}_{[a_k,b_k]}h(u), & u\in[a_k,b_k] \text{ for some } k \\ \mathcal{D}^{(1,0)}_{[a_k,b_k]}h(b_k)\left(\dfrac{a_{k+1}-u}{a_{k+1}-b_k}\right), & u\in(b_k,a_{k+1}] \text{ for some } k. \end{cases}$$

It follows that h^D is continuous and $h^D\geq 0$ so the redefined version of h^* also satisfies Properties C1-C3. Redefine $f_n^*(u) = t_n^{-1}\{\mathcal{M}(R+t_nh^*)-\mathcal{M}R\}(u)-h(u)$. Similar to before, we use Dini's theorem to conclude $\lim_{n\to\infty}\|f_n^*\|_\infty = \|h^*-h\|_\infty = \|h^D\|_\infty$ and

$$\begin{aligned}
d\mathcal{D}_R h(u) = d\mathcal{M}_R h(u) - h(u) &\leq t_n^{-1}\{\mathcal{M}(R+t_nh)-\mathcal{M}R\}(u)-h(u) \\
&\leq t_n^{-1}\{\mathcal{M}(R+t_nh^*)-\mathcal{M}R\}(u)-h(u) = f_n^*(u).
\end{aligned}$$

It follows that $\limsup_{n\to\infty}\|f_n\|_\infty \leq \max\{\|f\|_\infty, \|h^D\|_\infty\}$ and hence it suffices to show $\|f\|_\infty = \|d\mathcal{D}_R h\|_\infty \geq \|h^D\|_\infty$. If u is in one of the non-strictly star-shaped regions in $\cup_k[a_k,b_k]$, then $|h^D(u)| = h^D(u) = \mathcal{D}^{(1,0)}_{[a_k,b_k]}h(u) = d\mathcal{D}_R h(u) \leq \sup_{u\in\cup_k[a_k,b_k]}|d\mathcal{D}_R h(u)|$. If $u\in(\cup_k[a_k,b_k])^c$, then

$$|h^D(u)| = h^D(u) = \mathcal{D}^{(1,0)}_{[a_k,b_k]}h(b_k)\left(\frac{a_{k+1}-u}{a_{k+1}-b_k}\right) \leq \mathcal{D}^{(1,0)}_{[a_k,b_k]}h(b_k) \leq \sup_{u\in\cup_k[a_k,b_k]}|d\mathcal{D}_R h(u)|.$$

We conclude that $\|h^D\|_\infty \leq \|d\mathcal{D}_R h\|_\infty$ and thus complete the proof of Lemma A.6. $\square$

Remark A.7. *The next lemma generalizes Lemmas A.4 and A.6 to allow for ODCs $R \in \Theta_1$. This lemma is critical in proving Theorem 2.3 and in justifying subsequent results. We use the same notation as in Lemmas A.4 and A.6; note that the operator $\mathcal{L}_{S_k}$ and the S_k set notation are defined in Section 2.3.2 of the manuscript.*

Lemma A.7. *Suppose $R \in \Theta_1$. Suppose $h \in C([0,1])$, $h_n \in l([0,1])$, and $u \in [0,1]$. When $\|h_n - h\|_\infty \to 0$ and $t_n \downarrow 0$ as $n \to \infty$, then*

(a) Generalization of Lemma A.4:

$$\begin{aligned}
\lim_{n\to\infty} t_n^{-1} \left\{\mathcal{M}(R + t_n h_n) - \mathcal{M}R\right\}(u) &= d\mathcal{M}_R h(u) \\
\lim_{n\to\infty} t_n^{-1} \left\{\mathcal{D}(R + t_n h_n) - \mathcal{D}R\right\}(u) &= d\mathcal{D}_R h(u)
\end{aligned}$$

where

$$d\mathcal{M}_R h(u) = \begin{cases} 0, & \text{if } u = 1,\ R(u) < 1 \\ \max\{h(1), 0\}, & \text{if } u = 1,\ R(u) = 1 \\ \mathcal{L}_{S_k} h(u), & \text{if } \exists k \text{ such that } u \in S_k \setminus \{1\} \\ h(u), & \text{otherwise} \end{cases}$$

and $d\mathcal{D}_R h = d\mathcal{M}_R h - h$.

(b) Generalization of Lemma A.6: *For all $p \in [1, \infty]$,*

$$\begin{aligned}
\lim_{n\to\infty} \left\| t_n^{-1} \left\{\mathcal{M}(R + t_n h_n) - \mathcal{M}R\right\}\right\|_p &= \|d\mathcal{M}_R h\|_p \\
\lim_{n\to\infty} \left\| t_n^{-1} \left\{\mathcal{D}(R + t_n h_n) - \mathcal{D}R\right\}\right\|_p &= \|d\mathcal{D}_R h\|_p .
\end{aligned}$$

Proof of Lemma A.7. We first prove (a). When $R \in \Theta_1$, we need to account for the non-star-shaped regions in $\cup_k S_{k2}$. As in Lemma A.4, it suffices to show that $\lim_{n\to\infty} t_n^{-1}\{\mathcal{M}(R + t_n h)(u) - \mathcal{M}R(u)\} = d\mathcal{M}_R h(u)$ for $u \in [0,1]$. When $u = 1$ and $R(1) = 1$, it follows that $t_n^{-1}\{\mathcal{M}(R + t_n h)(1) - \mathcal{M}R(1)\} = \max\{h(1), 0\}$. If $R(1) < 1$, then for n sufficiently large, $R(1) + t_n h(1) < 1$ and $t_n^{-1}\{\mathcal{M}(R + t_n h)(1) -$

$\mathcal{M}R(1)\} = t_n^{-1}[\max\{R(1) + t_n h(1), 1\} - 1] = 0$. Thus, the boundary behaves $\lim_{n\to\infty} |t_n^{-1}\{\mathcal{M}(R + t_n h)(1) - \mathcal{M}R(1)\} - d\mathcal{M}_R h(1)| = 0$.

Suppose $u \in [0, 1)$. Let S denote the strictly star-shaped region of $\mathcal{M}R$ and $S_k = [a_k, b_k]$, $k = 1, 2, ...$, denote the non-strictly star-shaped regions of $\mathcal{M}R$. Recall that each $S_k = S_{k1} \cup S_{k2}$, where $S_{k1} = \{u \in S_k : \mathcal{M}R(u) = R(u)\}$ and $S_{k2} = \{u \in S_k : \mathcal{M}R(u) > R(u)\}$. Note that $\inf_u S_{k1} = a_k$. Let $S_0 = S \cup (\cup_k S_{k1})$ and observe that $\mathcal{M}R(v) = R(v)$ over S_0 and $\mathcal{M}R(v) > R(v)$ over $\cup_k S_{k2}$. Denote the secant line through $(u, \mathcal{M}R(u))$ and $(1, 1)$ by ξ_u and let $h_{n,u}^R = t_n^{-1}(R - \xi_u) + h$. Then $\mathcal{M}R(u) = \xi_u(u)$ and $t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\}(u) = \mathcal{M}_{[0,1]}^{(1,0)} h_{n,u}^R(u)$. As in the proof of Lemma A.4, we set $a_u = a_k$ when $u \in S_k$ and $a_u = u$ otherwise. Denote by $S(u) = \{v \in [0, u] : v \in [a_u, u] \cap S_0\}$ and $S(u)^c = \{v \in [0, u] : v \notin S(u)\}$. When $v \in S(u)$, $R(v) = \xi_u(v)$ and $h_{n,u}^R(v) = h(v)$. When $v \in S(u)^c$, $h_{n,u}^R(v) < h(v)$. Define

$$\psi(u) = -\inf_{v\in S(u)} \left\{\frac{-h(v)}{1 - v}\right\}(1 - u), \quad \text{for } u \in [0, 1).$$

It follows that

$$\psi(u) = -\inf_{v\in[a_k,u]\cap S_{k1}} \left\{\frac{-h(v)}{1 - v}\right\}(1 - u) = -\inf_{v\in S_{k1}, v\le u} \left\{\frac{-h(v)}{1 - v}\right\}(1 - u)$$

when $u \in [a_k, b_k]$ and $\psi(u) = h(u)$ otherwise. It suffices to show that for each $u \in [0, 1)$, $\lim_{n\to\infty} |\mathcal{M}_{[0,1]}^{(1,0)} h_{n,u}^R(u) - \psi(u)| = 0$.

If $a_u = 0$, then this holds trivially. If $a_u > 0$, suppose $\delta > 0$. Applying the triangle inequality yields

$$\left|\mathcal{M}_{[0,1]}^{(1,0)} h_{n,u}^R(u) - \psi(u)\right| \le \left|\mathcal{M}_{[0,1]}^{(1,0)} h_{n,u}^R(u) - \psi_{\delta,n}(u)\right| + |\psi_{\delta,n}(u) - \psi(u)|, \qquad \text{(A.10)}$$

where

$$\psi_{\delta,n}(u) = -\inf_{v\in S_\delta(u)} \left\{\frac{-h_{n,u}^R(v)}{1 - v}\right\}(1 - u)$$

and $S_\delta(u) = \{w \in [0, u] : \inf_{v\in S(u)} |w - v| < \delta\}$ is a δ-enlargement set of $S(u)$ for $u \in [0, 1)$. We consider the two terms on the right-hand side of (A.10) separately. We first show $\lim_{n\to\infty} |\mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u) - \psi_{\delta,n}(u)| = 0$. Let $S_\delta(u)^c = \{w \in [0,u] : w \notin S_\delta(u)\}$, a closed subset of $S(u)^c$. When δ is sufficiently small, $S_\delta(u)^c$ is non-empty. For each $v \in S_\delta(u)^c$, $-h^R_{n,u}(v)/(1-v)$ diverges, and thus there exists an $N_1(u,v,\delta) \geq 1$ such that for all $n \geq N_1(u,v,\delta)$,

$$\frac{-h^R_{n,u}(v)}{1-v} > \inf_{v\in S(u)} \left\{ \frac{-h(v)}{1-v} \right\} = \inf_{v\in S(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\}.$$

Define

$$E_n = \left\{ v \in S(u)^c : \frac{-h^R_{n,u}(v)}{1-v} > \inf_{v\in S(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} \right\}.$$

This sequence of sets forms an open covering, relative to $[0,1]$, of $S_\delta(u)^c$, which is compact. Thus, there exists a finite number of E_n sets whose union covers $S_\delta(u)^c$; i.e., there exists an $N_2(u,\delta) \geq 1$ such that

$$\inf_{v\in S_\delta(u)^c} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} \geq \inf_{v\in S(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\}$$

and hence

$$\begin{aligned} &\inf_{v\in[0,u]} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} \\ &\quad = \min\left\{ \inf_{v\in S(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\}, \inf_{v\in D(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\}, \inf_{v\in S_\delta(u)^c} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} \right\} \\ &\quad = \min\left\{ \inf_{v\in S(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\}, \inf_{v\in D(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} \right\} = \inf_{v\in S_\delta(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} \end{aligned}$$

for all $n \geq N_2(u,\delta)$, where $D(u) = S(u)^c \cap S_\delta(u)$. Therefore, we have

$$\begin{aligned} &\lim_{n\to\infty} \left| \mathcal{M}^{(1,0)}_{[0,1]} h^R_{n,u}(u) - \psi_{\delta,n}(u) \right| \\ &\quad = \lim_{n\to\infty} \left| \inf_{v\in S_\delta(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} (1-u) - \inf_{v\in[0,u]} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} (1-u) \right| = 0. \end{aligned}$$

We now show $|\psi_{\delta,n}(u) - \psi(u)|$ converges to 0 uniformly in n as $\delta \to 0$. Note that $S(u) \subset S_\delta(u)$, so write $S_\delta(u) = S(u) \cup D(u)$. It follows that

$$\begin{aligned} \psi_{\delta,n}(u) &= -\inf_{v\in S_\delta(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} (1-u) \\ &= -\min\left\{ \inf_{v\in S(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\}, \inf_{v\in D(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} \right\} (1-u) \\ &\geq -\inf_{v\in S(u)} \left\{ \frac{-h^R_{n,u}(v)}{1-v} \right\} (1-u) = -\inf_{v\in S(u)} \left\{ \frac{-h(v)}{1-v} \right\} (1-u) = \psi(u). \end{aligned}$$

Since $-t_n^{-1}\{R(v)-\xi_u(v)\} > 0$ when $v \in D(u)$ and $-t_n^{-1}\{R(v)-\xi_u(v)\} = 0$ when $v \in S(u)$, we also have

$$\begin{aligned}
\psi_{\delta,n}(u) &= -\inf_{v\in S_\delta(u)} \left\{\frac{-h^R_{n,u}(v)}{1-v}\right\}(1-u) \\
&= -\inf_{v\in S_\delta(u)} \left[\frac{-t_n^{-1}\{R(v)-\xi_u(v)\}-h(v)}{1-v}\right](1-u) \\
&= -\min\left[\inf_{v\in S(u)}\left\{\frac{-h(v)}{1-v}\right\}, \inf_{v\in D(u)}\left[\frac{-t_n^{-1}\{R(v)-\xi_u(v)\}-h(v)}{1-v}\right]\right](1-u) \\
&\le -\min\left[\inf_{v\in S(u)}\left\{\frac{-h(v)}{1-v}\right\}, \inf_{v\in D(u)}\left\{\frac{-h(v)}{1-v}\right\}\right](1-u) \\
&= -\inf_{v\in S_\delta(u)}\left\{\frac{-h(v)}{1-v}\right\}(1-u) \equiv \psi^*_\delta(u),
\end{aligned}$$

which is free of n. We have shown that $\psi(u) \le \psi_{\delta,n}(u) \le \psi^*_\delta(u)$ and thus $|\psi_{\delta,n}(u) - \psi(u)| \le \psi^*_\delta(u) - \psi(u)$. Therefore, to establish part (a), it is sufficient to show that $\lim_{\delta\to 0}\{\psi^*_\delta(u) - \psi(u)\} = 0$. The difference $\psi^*_\delta(u) - \psi(u)$ can be written as

$$\sup_{v\in S_\delta(u)} H_u(v) - \sup_{v\in S(u)} H_u(v) = \max\left\{\sup_{v\in D(u)} H_u(v), \sup_{v\in S(u)} H_u(v)\right\} - \sup_{v\in S(u)} H_u(v)$$

where the function $H_u(v)$ is defined in Lemma A.4 for $0 \le v \le u < 1$. Since $H_u(v)$ is uniformly continuous over $[0,u]$, then for any $\epsilon > 0$ there exists a $\delta(\epsilon,u) > 0$ such that $|H_u(v) - H_u(w)| < \epsilon$ whenever $|v-w| < \delta(\epsilon,u)$. Suppose $0 < \delta < \delta(\epsilon,u)$. According to the definition of $S_\delta(u)$, for any $v \in D(u)$, there exists $w_v \in S(u)$ such that $|v - w_v| < \delta$ and thus $H_u(v) < H_u(w_v) + \epsilon$. Therefore, $\sup_{v\in D(u)} H_u(v) \le \sup_{v\in D(u)} H_u(w_v) + \epsilon \le \sup_{v\in S(u)} H_u(v) + \epsilon$ and

$$\begin{aligned}
0 &\le \max\left\{\sup_{v\in D(u)} H_u(v), \sup_{v\in S(u)} H_u(v)\right\} - \sup_{v\in S(u)} H_u(v) \\
&\le \max\left\{\sup_{v\in S(u)} H_u(v)+\epsilon, \sup_{v\in S(u)} H_u(v)\right\} - \sup_{v\in S(u)} H_u(v) \\
&\le \sup_{v\in S(u)} H_u(v) + \epsilon - \sup_{v\in S(u)} H_u(v) = \epsilon.
\end{aligned}$$

This proves part (a).

We now prove part (b). When $1 \le p < \infty$, part (b) follows directly from the monotone convergence theorem. When $p = \infty$, let

$$f_n(u) = t_n^{-1}\{\mathcal{M}(R + t_n h) - \mathcal{M}R\}(u)$$

and $f(u) = d\mathcal{M}_R h(u)$ for $u \in [0,1]$. Our first goal is to show that (A.8) holds for $R \in \Theta_1$; i.e., $\lim_{n\to\infty} \|f_n\|_\infty = \|f\|_\infty$. Showing $\liminf_{n\to\infty} \|f_n\|_\infty \ge \|f\|_\infty$ uses the same argument described in Lemma A.6. However, the approach we used in proving Lemma A.6 (for $R \in \Theta_0$) cannot be used to establish $\limsup_{n\to\infty} \|f_n\|_\infty \le \|f\|_\infty$ when $R \in \Theta_1$ because for certain functions $h \in l([0,1])$ there is no h^* that satisfies Properties C1-C3. Therefore, a different type of argument is needed.

Suppose $f \ge 0$. For each $u \in [0,1]$, $f_n(u)$ is nonincreasing and converges to $f(u)$ as $n \to \infty$. Therefore, $f_n \ge 0$ and $\|f_n\|_\infty$ is bounded and nonincreasing and thus $\lim_{n\to\infty} \|f_n\|_\infty$ exists; i.e., $\lim_{n\to\infty} \|f_n\|_\infty = \liminf_{n\to\infty} \|f_n\|_\infty = \limsup_{n\to\infty} \|f_n\|_\infty$. Since $f_n \ge 0$ is continuous over $[0,1]$, there exists an $x_n \in [0,1]$ such that $f_n(x_n) = \|f_n\|_\infty$. Since the sequence $\{x_n\}_{n=1}^\infty$ is bounded, we can find a subsequence $\{x_{k_n}\}_{n=1}^\infty$ such that $\lim_{n\to\infty} x_{k_n} = x_0 \in [0,1]$. For each k_n, whenever $k_m > k_n$, $f_{k_n}(x_{k_m}) \ge f_{k_m}(x_{k_m}) \ge \lim_{n\to\infty} \|f_n\|_\infty$. Thus, $\lim_{m\to\infty} f_{k_n}(x_{k_m}) = f_{k_n}(x_0) \ge \lim_{n\to\infty} \|f_n\|_\infty$. Consequently, $f(x_0) = \lim_{n\to\infty} f_{k_n}(x_0) \ge \lim_{n\to\infty} \|f_n\|_\infty$ which implies $\|f\|_\infty \ge \lim_{n\to\infty} \|f_n\|_\infty$ and thus $\lim_{n\to\infty} \|f_n\|_\infty = \|f\|_\infty$ holds when $f \ge 0$.

For general f, write $f = f^+ + f^-$ and $f_n = f_n^+ + f_n^-$, where $f^+ = \max\{f, 0\}$, $f^- = \min\{f, 0\}$, and similarly for f_n^+ and f_n^-. Both f_n^+ and f_n^- are continuous over $[0,1]$. Furthermore, for each $u \in [0,1]$, both $f_n^+(u)$ and $f_n^-(u)$ are nonincreasing sequences that converge to $f^+(u)$ and $f^-(u)$, respectively. Because $f^- \le f_n^- \le 0$, it follows that $\|f^-\|_\infty = \sup_{u\in[0,1]}\{-f^-(u)\} \ge \sup_{u\in[0,1]}\{-f_n^-(u)\} = \|f_n^-\|_\infty$. Therefore, $\|f^-\|_\infty \ge \limsup_{n\to\infty} \|f_n^-\|_\infty$ and hence $\lim_{n\to\infty} \|f_n^-\|_\infty = \|f^-\|_\infty$. Since $f^+ \ge 0$, the argument in the last paragraph shows that $\lim_{n\to\infty} \|f_n^+\|_\infty = \|f^+\|_\infty$. That $\lim_{n\to\infty} \|f_n\|_\infty = \|f\|_\infty$ follows from the fact $\|f_n\|_\infty = \max\{\|f_n^+\|_\infty, \|f_n^-\|_\infty\}$ and $\|f\|_\infty = \max\{\|f^+\|_\infty, \|f^-\|_\infty\}$. Then (A.8) holds for $R \in \Theta_1$ and $p = \infty$.

To show that (A.9) holds for $R \in \Theta_1$ and $p = \infty$, redefine f_n and f as $f_n(u) = t_n^{-1}\{\mathcal{M}(R+t_n h) - \mathcal{M}R\}(u) - h(u)$ and $f(u) = d\mathcal{D}_R h(u)$ for $u \in [0,1]$. Observe that f is bounded on $[0,1]$. Also, the sequence of functions f_n is nonincreasing, continuous over $[0,1]$, and converges pointwise to f as $n \to \infty$. The same argument used to prove (A.8) for $R \in \Theta_1$ and $p = \infty$ now applies. □

Lemma A.8. *Suppose $R \in \Theta_0$ is non-strictly star-shaped over closed, pairwise disjoint intervals of the form $[a_k, b_k]$, $0 \le a_k < b_k \le 1$, as described in Section 2.3.1. For each k, define the index set $T_k = \{(u,v) : a_k \le v \le u < b_k\}$ and*

$$Q_k(u,v) = \left(\frac{1-u}{1-v}\right) T_R^\lambda(v) - T_R^\lambda(u),$$

where $(u,v) \in T_k$ and where T_R^λ is the process defined in the proof of Theorem 2.1. Then, the processes $\{Q_k(u,v); (u,v) \in T_k\}$ are mutually independent among k.

Proof of Lemma A.8. Let $\mathbf{t}_{kn} = \{(u_{k1}, v_{k1}), \ldots, (u_{kn}, v_{kn})\}$ denote a finite collection of points in T_k, for $k = 1, 2, \ldots$. Let $\mathbf{Q}_k(\mathbf{t}_{kn}) = (Q_k(u_{k1}, v_{k1}), \ldots, Q_k(u_{kn}, v_{kn}))'$. For any $k_1, \ldots, k_m$, it is easy to show that the random vector $(\mathbf{Q}'_{k_1}(\mathbf{t}_{k_1 n}), \ldots, \mathbf{Q}'_{k_m}(\mathbf{t}_{k_m n}))'$ is multivariate normal. It therefore suffices to show that $\mathbf{Q}_{k_i}(\mathbf{t}_{k_i n})$ and $\mathbf{Q}_{k_j}(\mathbf{t}_{k_j n})$ are uncorrelated for any $1 \le i < j \le m$. It suffices to show $\mathrm{cov}(Q_1(u_{11}, v_{11}), Q_2(u_{21}, v_{21})) = 0$, where $a_1 \le v_{11} \le u_{11} \le b_1 < a_2 \le v_{21} \le u_{21} \le b_2$. This covariance equals

$$\begin{aligned}
&\mathrm{cov}\left(\left(\frac{1-u_{11}}{1-v_{11}}\right) T_R^\lambda(v_{11}) - T_R^\lambda(u_{11}), \left(\frac{1-u_{21}}{1-v_{21}}\right) T_R^\lambda(v_{21}) - T_R^\lambda(u_{21})\right) \\
&= \left(\frac{1-u_{11}}{1-v_{11}}\right)\left(\frac{1-u_{21}}{1-v_{21}}\right)\{\lambda R'(a_2)R(v_{11})(1-v_{21}) + (1-\lambda)R'(a_1)R'(a_2)v_{11}(1-v_{21})\} \\
&\quad - \left(\frac{1-u_{11}}{1-v_{11}}\right)\{\lambda R'(a_2)R(v_{11})(1-u_{21}) + (1-\lambda)R'(a_1)R'(a_2)v_{11}(1-u_{21})\} \\
&\quad - \left(\frac{1-u_{21}}{1-v_{21}}\right)\{\lambda R'(a_2)R(u_{11})(1-v_{21}) + (1-\lambda)R'(a_1)R'(a_2)u_{11}(1-v_{21})\} \\
&\qquad + \lambda R'(a_2)R(u_{11})(1-u_{21}) + (1-\lambda)R'(a_1)R'(a_2)u_{11}(1-u_{21}),
\end{aligned}$$

which reduces to 0. □

A.2 Densities and critical values

We provide the estimated probability density function of $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$ with different choices of p where $p \in \{1, 2, 3, 5, \infty\}$ in Figure A.1 and its upper $\alpha = 0.01$, 0.05, and 0.10 quantiles in Table A.1. For fixed $p \in [1, \infty]$, $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$ represents the distribution of M^p_{mn} under the least favorable configuration in Θ_0, that is, when $F = G$, as shown in Theorem 2.2 in Chapter 2.

Figure A.1 and Table A.1 were created using the simulation techniques described at the end of Section 2.3.1 of the manuscript.

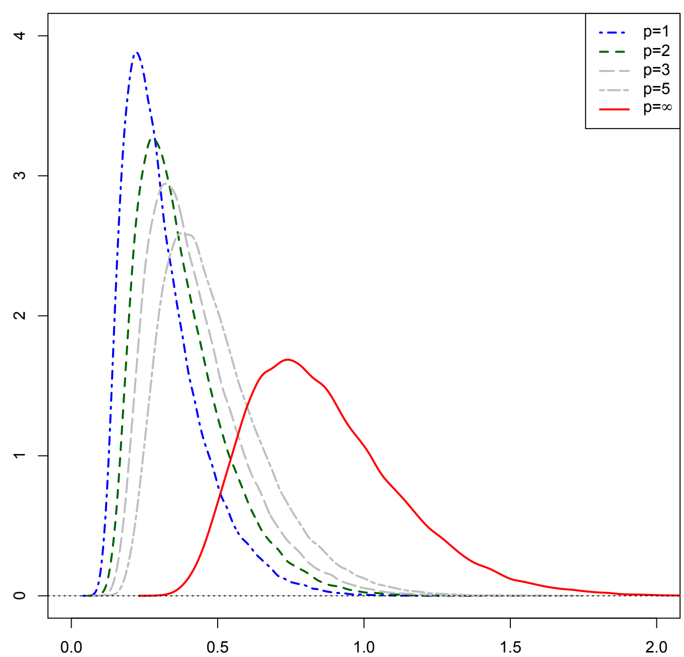

Figure A.1: Probability density function of $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$ for $p \in \{1, 2, 3, 5, \infty\}$.

Table A.1: Values of $c_{\alpha,p}$, the upper α quantiles of $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$, for $p \in \{1,2,3,5,\infty\}$ and $\alpha = 0.01$, 0.05, and 0.10.

α	$p=1$	$p=2$	$p=3$	$p=5$	$p=\infty$
0.01	0.751	0.860	0.939	1.047	1.623
0.05	0.580	0.676	0.746	0.841	1.353
0.10	0.496	0.586	0.651	0.738	1.219

A.3 Supplementary material for Section 2.3.2

The following excerpt is taken from Section 2.3.2:

> To characterize a non-star-shaped ODC $R \in \Theta_1$, start with $\mathcal{M}R$, which is star-shaped, and note that (as in Section 2.3.1) one can partition the unit interval $[0,1]$ as $[0,1] = S \cup (\cup_k S_k)$, where $\mathcal{M}R$ is strictly star-shaped over S and non-strictly star-shaped over pairwise disjoint intervals of the form $S_k = [a_k, b_k]$, $0 \le a_k < b_k \le 1$, for $k = 1, 2, \ldots$. One can further partition each S_k as $S_k = S_{k1} \cup S_{k2}$, where $S_{k1} = \{u \in S_k : \mathcal{M}R(u) = R(u)\}$ and $S_{k2} = \{u \in S_k : \mathcal{M}R(u) > R(u)\}$. Each S_{k1} must contain a_k so it is never empty, and the non-star-shaped regions of R can be written as $\cup_k S_{k2}$. In other words, $R \in \Theta_0$ when $\cup_k S_{k2}$ is empty and $R \in \Theta_1$ otherwise.

With this notation in mind, we introduced the functional operator $\mathcal{L}_{S_k} : l([0,1]) \mapsto l([0,1])$ according to

$$\mathcal{L}_{S_k} h(u) = - \inf_{\substack{v \in S_{k1} \\ v \le u}} \left\{ \frac{-h(v)}{1-v} \right\} (1-u) I_{S_k}(u) + h(u) I_{S_k^c}(u), \quad \text{for } u \in [0,1),$$

with suitable modification at $u = 1$ similar to that for $\mathcal{M}^{(c,d)}_E h(c)$ in Lemma A.1. This operator plays a key role in describing the limiting distribution of $c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p$ for $p \in [1,\infty]$ and for a general $R \in \Theta$; see Theorem 2.3 in Chapter 2. We provide an illustrative example.

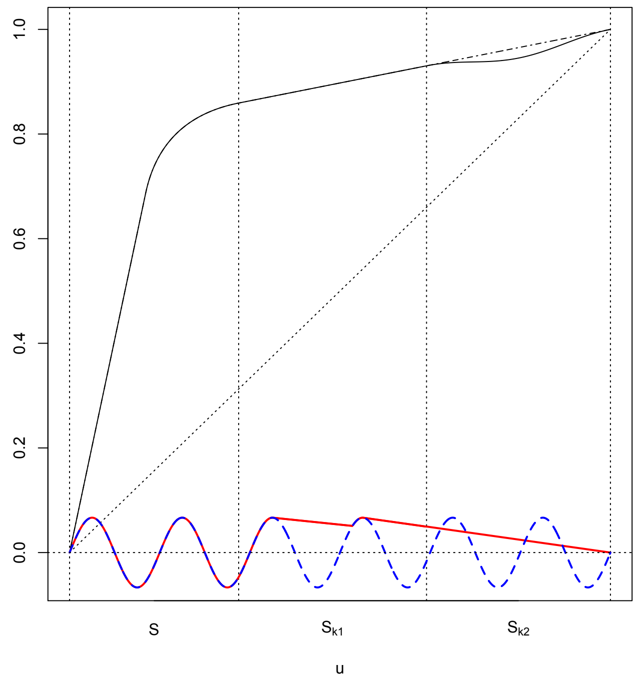

Figure A.2: Illustration of the functional operator $\mathcal{L}_{S_k}$ when $k = 1$.

Figure A.2 shows how $\mathcal{L}_{S_k} h(u)$ is calculated with $k = 1$; a generic function $h(u)$ is shown at the bottom of the figure in blue and the function $\mathcal{L}_{S_k} h(u)$ is shown in red. The solid curve is the true ODC R and the dot-dashed line is the least star-shaped majorant $\mathcal{M}R$. Note that $\mathcal{M}R$ is strictly star-shaped over S and non-strictly star-shaped over $S_k = S_{k1} \cup S_{k2}$, where $\mathcal{M}R(u) = R(u)$ when $u \in S_{k1}$ and $\mathcal{M}R(u) > R(u)$ when $u \in S_{k2}$. An arbitrary function h is shown at the bottom in blue and $\mathcal{L}_{S_k} h$ is shown in red. The equal distribution line $R(u) = u$ is shown dotted.

In Figure A.2, note that

- over the strictly star-shaped region S, $\mathcal{L}_{S_k} h(u) = h(u)$
- over the non-strictly star-shaped region S_{k1}, $\mathcal{L}_{S_k} h(u)$ is the least star-shaped majorant of $h(u)$ with kernel $(1, 0)$; i.e., $\mathcal{L}_{S_k} h(u) = \mathcal{M}_{S_k}^{(1,0)} h(u)$
- over the non-star-shaped region S_{k2}, $\mathcal{L}_{S_k} h(u)$ is calculated by taking $\mathcal{M}_{S_k}^{(1,0)} h(u)$ over S_{k1} and extending it linearly to the point $(1, 0)$. Note also that $\mathcal{L}_{S_k} h(u)$

does not depend on the function $h(u)$ over S_{k2} but only on $\mathcal{M}_{S_k}^{(1,0)}h(u)$ over S_{k1}.

From these observations, we now show that the limiting distributions in Theorem 2.3 in Chapter 2 (for general $R \in \Theta$) reduce to those in Theorem 2.1 when $R \in \Theta_0$.

Proof: Suppose $p \in [1, \infty)$. When $R \in \Theta_0$ is non-strictly star-shaped, the non-star-shaped region described by $\cup_k S_{k2}$ is empty. Therefore, for each k, $S_k = S_{k1} = [a_k, b_k]$, $\mathcal{L}_{S_k}$ reduces to $\mathcal{M}_{[a_k,b_k]}^{(1,0)}$, and $\mathcal{L}_{S_k} - \mathcal{I} = \mathcal{M}_{[a_k,b_k]}^{(1,0)} - \mathcal{I} = \mathcal{D}_{[a_k,b_k]}^{(1,0)}$. Since $\mathcal{D}R = \mathcal{M}R - R = 0$, the statement in Theorem 2.3

$$c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p \xrightarrow{d} \left\{\sum_k \int_{u\in S_k} \left|\mathcal{L}_{S_k}T_R^\lambda(u) - T_R^\lambda(u)\right|^p du\right\}^{1/p}$$

can be written as

$$c_{mn}\|\mathcal{D}R_{mn}\|_p \xrightarrow{d} \left[\sum_k \int_{a_k}^{b_k} \left\{\mathcal{D}_{[a_k,b_k]}^{(1,0)}T_R^\lambda(u)\right\}^p du\right]^{1/p}.$$

But, $M_{mn}^p = c_{mn}\|\mathcal{D}R_{mn}\|_p$ when $R \in \Theta_0$ and from the proof of Theorem 2.1, we have already shown that

$$\left[\sum_k \int_{a_k}^{b_k} \left\{\mathcal{D}_{[a_k,b_k]}^{(1,0)}T_R^\lambda(u)\right\}^p du\right]^{1/p}$$
$$\stackrel{d}{=} \left\{\sum_k \left[\lambda R'(a_k) + (1-\lambda)\{R'(a_k)\}^2\right]^{p/2} \int_{a_k}^{b_k} \left\{\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}(u)\right\}^p du\right\}^{1/p},$$

the limiting distribution identified in Theorem 2.1.

When $R \in \Theta_0$ is strictly star-shaped, then all of the S_k's are empty and $\mathcal{L}_{S_k}T_R^\lambda(u) = T_R^\lambda(u)$ for all $u \in [0, 1]$. Clearly, in this case, we have $M_{mn}^p \xrightarrow{d} 0$. The argument for the sup-norm case is completely analogous. □

A.4 Supplementary material for Section 2.4

In Sections 2.4.1 and 2.4.2, the ODCs R_1, R_2, R_3, and R_4 are star-shaped (i.e., $R_i \in \Theta_0$, for $i = 1, 2, 3, 4$) and are each members of $\{R(u; \delta_1, \delta_2) : 0 \leq u \leq 1; -1 < \delta_1 < 1, 0 \leq \delta_2 \leq 1\}$, a family of ODCs described by two parameters δ_1 and δ_2. This family is an extension of the family used in Beare and Moon (2015), who used Bézeir curves to construct concave ODCs using one parameter. Introducing a second parameter allows us to describe star-shaped ODCs (and also non-star-shaped ODCs). The formula for $R(u; \delta_1, \delta_2)$ is

$$R(u; \delta_1, \delta_2) =$$
$$\begin{cases} u, & 0 \leq u \leq u_1 \\ \{1 - t_1(u)\}^2 u_1 + 2\{1 - t_1(u)\} t_1(u) m_1 + t_1^2(u)\{m_1 + r(v_1 - m_1)\}, & u_1 < u < v_1 \\ m_1 + r(u - m_1), & v_1 \leq u \leq u_2 \\ \{1 - t_2(u)\}^2\{m_1 + r(u_2 - m_1)\} & \\ \qquad + 2\{1 - t_2(u)\} t_2(u)(r^{-1} m_2 + c) + t_2^2(u)(r^{-1} v_2 + c), & u_2 < u < v_2 \\ r^{-1} u + c, & v_2 \leq u \leq 1, \end{cases}$$

where $u_1 = \frac{3}{4}(1 - \delta_2)$, $u_2 = a + (1 - a)(1 - \delta_2)$,

$$\begin{aligned} v_1 &= \frac{3}{4}(1 - \delta_2) + \frac{1}{4}\{a + (1 - a)(1 - \delta_2)\} \\ v_2 &= b + (1 - b)(1 - \delta_2) \\ m_1 &= 1 - \delta_2 \\ m_2 &= \{r - 1 - m_1 r(1 - r)\}(r^2 - 1)^{-1} \\ r &= (1 + \delta_1)(1 - \delta_1)^{-1} \\ a &= 4(5r + 4)^{-1} \\ b &= r(5 - 4c)(5r + 4)^{-1} \\ c &= 2\delta_1(1 + \delta_1)^{-1} \end{aligned}$$

and

$$t_i(u) = \frac{u_i - m_i + \{m_i^2 - u_i v_i + (u_i + v_i - 2m_i)u\}^{1/2}}{u_i + v_i - 2m_i},$$

for $i = 1, 2$ and $0 \leq u \leq 1$. The four H_0 ODCs used in the manuscript are

- $R_1(u) = R(u; 0.4, 0.4)$
- $R_2(u) = R(u; 0.4, 0.8)$
- $R_3(u) = R(u; 0.8, 0.4)$
- $R_4(u) = R(u; 0.8, 0.8)$.

The four H_1 ODCs R_5, R_6, R_7, and R_8 represent a gallimaufry of curves described below:

- $R_5(u)$, a member of the family of ODCs to be described next
- $R_6(u) = R(u; -0.4, 0.8)$, a member of the family of ODCs described above
- $R_7(u) = FG^{-1}(u)$, where $F \sim \mathcal{N}(0, 1)$ and $G \sim \mathcal{N}(0, 1/4)$
- $R_8(u) = FG^{-1}(u)$, where $F \sim \mathcal{N}(0, 1)$ and $G \sim \mathcal{N}(0, 4)$.

Table A.2 (below) and Table A.3 (next page) provide the estimated type I error and power results from our simulation study in Sections 2.4.1 and 2.4.2, respectively. Table A.2 summarizes the two-sample results; Table A.3 summarizes the one-sample results. Both of these tables are referenced in Chapter 2.

Table A.2: Estimated probability of rejecting $H_0 : F \leq_{\text{US}} G$ for $p \in \{1, 2, \infty\}$, different sample size configurations, and $\alpha = 0.05$. All estimates are based on 10,000 Monte Carlo data sets. *ODC*s R_1, R_2, R_3, and R_4 satisfy H_0. *ODC*s R_5, R_6, R_7, and R_8 satisfy H_1.

ODC	m	20	50	50	100	200	500	1000
	n	20	50	100	100	500	500	1000
	$p = 1$	0.082	0.064	0.057	0.059	0.050	0.048	0.050
$F = G$	$p = 2$	0.077	0.060	0.054	0.055	0.048	0.048	0.048
	$p = \infty$	0.044	0.039	0.039	0.041	0.040	0.043	0.045
	$p = 1$	0.047	0.024	0.014	0.016	0.009	0.009	0.005
R_1	$p = 2$	0.051	0.032	0.025	0.025	0.019	0.019	0.014
	$p = \infty$	0.038	0.031	0.030	0.031	0.033	0.033	0.029
	$p = 1$	0.002	0.000	0.000	0.000	0.000	0.000	0.000
R_2	$p = 2$	0.003	0.001	0.000	0.000	0.000	0.000	0.000
	$p = \infty$	0.003	0.002	0.001	0.001	0.000	0.001	0.001
	$p = 1$	0.026	0.013	0.006	0.010	0.004	0.005	0.003
R_3	$p = 2$	0.038	0.026	0.016	0.021	0.015	0.017	0.011
	$p = \infty$	0.034	0.031	0.022	0.029	0.029	0.031	0.026
	$p = 1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R_4	$p = 2$	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	$p = \infty$	0.002	0.001	0.000	0.000	0.001	0.001	0.001
	$p = 1$	0.009	0.021	0.029	0.061	0.612	0.980	1.000
R_5	$p = 2$	0.011	0.027	0.046	0.098	0.806	0.999	1.000
	$p = \infty$	0.009	0.023	0.043	0.095	0.859	1.000	1.000
	$p = 1$	0.579	0.886	0.941	0.993	1.000	1.000	1.000
R_6	$p = 2$	0.594	0.912	0.958	0.996	1.000	1.000	1.000
	$p = \infty$	0.501	0.899	0.956	0.998	1.000	1.000	1.000
	$p = 1$	0.204	0.493	0.617	0.816	0.998	1.000	1.000
R_7	$p = 2$	0.209	0.518	0.642	0.848	1.000	1.000	1.000
	$p = \infty$	0.121	0.383	0.523	0.767	0.999	1.000	1.000
	$p = 1$	0.100	0.112	0.097	0.194	0.679	0.958	1.000
R_8	$p = 2$	0.129	0.211	0.229	0.424	0.969	1.000	1.000
	$p = \infty$	0.119	0.280	0.346	0.565	0.997	1.000	1.000

Table A.3: Estimated probability of rejecting $H_0 : F \leq_{\text{US}} G$, G known, for $p \in \{1, 2, \infty\}$ and $\alpha = 0.05$. The GOF test from Arcones and Samaniego (2000), *AS*, is included.

ODC	m	20	50	75	100	125	150	200
	$p=1$	0.039	0.041	0.042	0.043	0.043	0.046	0.047
$F=G$	$p=2$	0.037	0.041	0.042	0.045	0.043	0.045	0.047
	$p=\infty$	0.052	0.049	0.048	0.051	0.053	0.051	0.052
	AS	0.011	0.022	0.024	0.030	0.032	0.036	0.037
	$p=1$	0.004	0.007	0.006	0.006	0.004	0.005	0.004
R_1	$p=2$	0.009	0.014	0.015	0.016	0.014	0.015	0.014
	$p=\infty$	0.027	0.033	0.035	0.034	0.033	0.034	0.034
	AS	0.005	0.014	0.018	0.022	0.023	0.024	0.025
	$p=1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R_2	$p=2$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$p=\infty$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	AS	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$p=1$	0.001	0.002	0.002	0.002	0.003	0.002	0.002
R_3	$p=2$	0.008	0.012	0.010	0.009	0.012	0.010	0.012
	$p=\infty$	0.028	0.030	0.026	0.029	0.028	0.030	0.031
	AS	0.005	0.015	0.016	0.018	0.020	0.021	0.026
	$p=1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R_4	$p=2$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$p=\infty$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	AS	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$p=1$	0.010	0.070	0.168	0.316	0.480	0.639	0.853
R_5	$p=2$	0.019	0.120	0.294	0.499	0.682	0.825	0.958
	$p=\infty$	0.060	0.239	0.471	0.674	0.829	0.920	0.988
	AS	0.002	0.004	0.013	0.039	0.106	0.208	0.502
	$p=1$	0.571	0.909	0.980	0.997	0.999	1.000	1.000
R_6	$p=2$	0.585	0.921	0.985	0.998	1.000	1.000	1.000
	$p=\infty$	0.631	0.930	0.986	0.998	0.999	1.000	1.000
	AS	0.364	0.840	0.961	0.993	0.998	1.000	1.000
	$p=1$	0.388	0.751	0.903	0.971	0.990	0.997	0.999
R_7	$p=2$	0.391	0.775	0.924	0.981	0.995	0.999	1.000
	$p=\infty$	0.441	0.800	0.936	0.985	0.997	0.999	1.000
	AS	0.166	0.594	0.833	0.945	0.984	0.995	1.000
	$p=1$	0.003	0.045	0.146	0.329	0.550	0.748	0.952
R_8	$p=2$	0.022	0.278	0.615	0.864	0.973	0.994	1.000
	$p=\infty$	0.095	0.586	0.881	0.979	0.999	1.000	1.000
	AS	0.027	0.496	0.853	0.977	0.999	1.000	1.000

We now describe the family of ODCs used in our local power analysis in Section 2.4.3. This family is described by a single parameter $\delta \in [0, 0.5]$. Specific members of this family are denoted by $R_{(\delta)} = R_{(\delta)}(u)$. The formula for $R_{(\delta)}(u)$ is

$$R_{(\delta)}(u) = \begin{cases} u, & 0 \leq u < u_1 \\ \overline{u}_1\{1 - t_1(u)\}^2 + 2\overline{m}_1 t_1(u)\{1 - t_1(u)\} + \overline{v}_1\{t_1(u)\}^2, & u_1 \leq u < v_1 \\ \delta + 8(u - \delta), & v_1 \leq u < u_2 \\ \overline{u}_2\{1 - t_2(u)\}^2 + 2\overline{m}_2 t_2(u)\{1 - t_2(u)\} + \overline{v}_2\{t_2(u)\}^2, & u_2 \leq u < v_2 \\ \frac{8}{18} + \delta + \frac{1}{8}\{u - (\frac{1}{18} + \delta)\}, & v_2 \leq u < u_3 \\ \overline{u}_3\{1 - t_3(u)\}^2 + 2\overline{m}_3 t_3(u)\{1 - t_3(u)\} + \overline{v}_3\{t_3(u)\}^2, & u_3 \leq u < v_3 \\ u, & v_3 \leq u < 1, \end{cases}$$

where

$$u_1 = \frac{7\delta}{8}, \quad u_2 = \delta + \frac{7}{144}, \quad u_3 = \delta + \frac{4}{9}, \quad \overline{u}_1 = \frac{7}{8}\delta, \quad \overline{u}_2 = \delta + \frac{7}{18}, \quad \overline{u}_3 = \delta + \frac{71}{144},$$
$$v_1 = \delta + \frac{1}{144}, \quad v_2 = \delta + \frac{1}{9}, \quad v_3 = \frac{7\delta}{8} + \frac{9}{16}, \quad \overline{v}_1 = \delta + \frac{1}{18}, \quad \overline{v}_2 = \delta + \frac{65}{144}, \quad \overline{v}_3 = \frac{7\delta}{8} + \frac{9}{16},$$
$$m_1 = \delta, \quad m_2 = \delta + \frac{1}{18}, \quad m_3 = \delta + \frac{1}{2}, \quad \overline{m}_1 = \delta, \quad \overline{m}_2 = \delta + \frac{8}{18}, \quad \overline{m}_3 = \delta + \frac{1}{2},$$

and

$$t_i(u) = \frac{u_i - m_i + \{m_i^2 - u_i v_i + (u_i + v_i - 2m_i)u\}^{1/2}}{u_i + v_i - 2m_i},$$

for $i = 1, 2, 3$ and $0 \leq u \leq 1$.

Notes:

- The $\delta = 0$, 0.1, 0.2, 0.3, 0.4, and 0.5 members of this family are shown in Figure 2.4 in Chapter 2.
- In Section 2.4.3, we used sequences of ODCs $\{R^{(r)}, r = 1, 2, \dots\}$ that converge to Θ_0 at different rates (denoted by ζ_r).
  1. For $\zeta_r = \log r$, we used the $\delta = 0.5 - \frac{1}{2}(\log 50/\log r)$ members of this family.

2. For $\zeta_r = r^{2/5}$, we used the $\delta = 0.5 - \frac{1}{2}(50/r)^{2/5}$ members of this family.

3. For $\zeta_r = r^{1/2}$, we used the $\delta = 0.5 - \frac{1}{2}(50/r)^{1/2}$ members of this family.

- The ODC $R_5(u)$ used in Sections 2.4.1 and 2.4.2 (fixed sample size comparisons) is the $\delta = 0.25$ member of this family.

Finally, we present the sample size determination results using our approach outlined in Section 2.4.1. Complete results are shown in Table A.4.

Table A.4: Minimum sample sizes to detect specific departures from H_0 using $\alpha = 0.05$ and power $1 - \beta = 0.8$ for $R_i \in \Theta_1$.

R_i	$p = 1$	$p = 2$	$p = \infty$	$p = 1$	$p = 2$	$p = \infty$	AS	$p = 1$	$p = 2$	$p = \infty$
R_5	345	275	260	184	146	120	297	634	461	582
R_6	38	36	40	24	21	20	32	121	103	156
R_7	123	64	47	51	27	20	26	198	178	2295
R_8	36	33	41	21	20	21	31	483	368	578

- The leftmost section of Table A.4 displays the results with two samples using direct simulation; i.e., we used Monte Carlo simulation at each value of $m = n$ (starting at $m = n = 10$) and increased this sample size until we attained an 80 percent rejection rate of H_0.

- The middle section of Table A.4 is the same as the leftmost section except that it considers one-sample tests (with sample size m and G known). The AS (2000) test from Arcones and Samaniego (2000) is also shown.

- The rightmost section of Table A.4 shows our conservative sample size determination approach described in Section 2.4.1. Comparing this section with the leftmost section, it is clear that our approach is very conservative! Section 2.6 describes approaches one might use to reduce the conservatism. R code to perform direct simulation is available from the author.

Appendix B

Supplementary materials for Chapter 3

B.1 Proofs

We want to show that EL_1, EL_2, EL_3, and EL_4 are marginal distribution-free; i.e., that the test statistics do not depend on marginal distributions. Consider $U_i = F(X_i)$ and $V_i = G(Y_i)$ for $i = 1, \ldots, n$; according to definition, (U_i, V_i) has joint distribution C. Further, denote $B_{11} = [0, F(x)] \times [0, G(y)]$, $B_{12} = [0, F(x)] \times [G(y), 1]$, $B_{21} = [F(x), 1] \times [0, G(y)]$, and $B_{22} = [F(x), 1] \times [G(y), 1]$. Given $(x, y)' \in \mathbb{R}^2$ it can be proved that

$$P_n(A_{kl}) = P_n^0(B_{kl}),$$

for $1 \leq k, l \leq 2$, where P_n^0 is the probability measure corresponding to the transformed data $\{(U_i, V_i)\}_{i=1}^n$; i.e., $P_n^0(B_{ij})$ is the proportion of the transformed sample falling in region B_{ij}. Because the local test statistics are all constructed by P_n and the A_{kl}'s, it is equivalent to consider P_n^0 and B_{ij} such that $-2\ln R_{01}(x, y) = -2\ln R_{01}^0(F(x), G(y))$ and $-2\ln R_{12}(x, y) = -2\ln R_{12}^0(F(x), G(y))$, where R_{01}^0 and R_{12}^0 are local likelihood ratios with P_n replaced by P_n^0 in R_{01}. In summary, we rewrite the test statistics as

$$\begin{aligned} \text{EL}_1 &= -\frac{2}{n}\sum_{i=1}^n \ln R_{01}(U_i, V_i) & \text{EL}_2 &= -\frac{2}{n^2}\sum_{i=1}^n\sum_{j=1}^n \ln R_{01}(U_i, V_j) \\ \text{EL}_3 &= -\frac{2}{n}\sum_{i=1}^n \ln R_{12}(U_i, V_i) & \text{EL}_4 &= -\frac{2}{n^2}\sum_{i=1}^n\sum_{j=1}^n \ln R_{12}(U_i, V_j), \end{aligned}$$

where the U_i's and V_j's are all from uniform(0,1) which are marginal-free.

Proof of Proposition 3.2. From the previous discussion, the test statistics EL_1 and

EL_2 can be rewritten as

$$\mathrm{EL}_1 = -\frac{2}{n}\sum_{i=1}^{n} \ln R_{01}(U_i, V_i) \quad \text{and} \quad \mathrm{EL}_2 = -\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \ln R_{01}(U_i, V_j).$$

Denote by Π the independence copula; i.e., $\Pi(u, v) = uv$ for all $u, v \in [0, 1]$. We can see that EL_1 depends on the distributions of R_{01} and (U_i, V_i) which are both C. Similarly, EL_2 depends on the distributions of R_{01} and (U_i, V_j). When $i \neq j$, it is clear that U_i and V_j are independently identically distributed uniform $(0, 1)$, so that the joint distribution of (U_i, V_j) is the independence copula Π. In other words, EL_2 depends on C and Π. Under $\mathcal{H}_0$, $C = \Pi$. So we have shown that EL_1 and EL_2 only depend on Π which are distribution-free in finite-sample cases. □

B.2 Supremum of empirical likelihoods under independence restriction

To find the supremum of the empirical likelihood under an independent assumption, i.e., $\sup\{L(\tilde{H}) : H(x, y) = F(x)G(y)\}$, we define

$$\begin{aligned}
\mathcal{L}(\mathbf{p}) = \frac{1}{n}&\left\{\sum_{i:(X_i,Y_i)\in A_{11}} \ln p_i + \sum_{j:(X_j,Y_j)\in A_{12}} \ln p_j + \sum_{k:(X_k,Y_k)\in A_{21}} \ln p_k + \sum_{l:(X_l,Y_l)\in A_{22}} \ln p_l\right\} \\
&+ \lambda_1\left\{\sum_{i:(X_i,Y_i)\in A_{11}} p_i + \sum_{j:(X_j,Y_j)\in A_{12}} p_j + \sum_{k:(X_k,Y_k)\in A_{21}} p_k + \sum_{l:(X_l,Y_l)\in A_{22}} p_l - 1\right\} \\
&+ \lambda_2\left\{\sum_{j:(X_j,Y_j)\in A_{12}} p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k - \sum_{i:(X_i,Y_i)\in A_{11}} p_i \sum_{l:(X_l,Y_l)\in A_{22}} p_l\right\}.
\end{aligned}$$

To maximize $\mathcal{L}(\mathbf{p})$, take the partial derivative respect to p_i for i such that $(X_i, Y_i) \in A_{11}$, p_j for j such that $(X_j, Y_j) \in A_{12}$, p_k for k such that $(X_k, Y_k) \in A_{21}$, and p_l for l

such that $(X_l, Y_l) \in A_{22}$. Then we obtain

$$\begin{aligned}
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_i} &= -\frac{1}{np_i} + \lambda_1 - \lambda_2 \sum_{l:(X_l,Y_l)\in A_{22}} p_l \\
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_j} &= -\frac{1}{np_j} + \lambda_1 + \lambda_2 \sum_{k:(X_k,Y_k)\in A_{21}} p_k \\
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_k} &= -\frac{1}{np_k} + \lambda_1 + \lambda_2 \sum_{j:(X_j,Y_j)\in A_{12}} p_j \\
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_l} &= -\frac{1}{np_l} + \lambda_1 - \lambda_2 \sum_{i:(X_i,Y_i)\in A_{11}} p_i.
\end{aligned}$$

Setting all partial derivatives equal to zero and solving, we obtain

$$\begin{aligned}
\frac{1}{n} &= \lambda_1 p_i - \lambda_2 p_i \sum_{i\in A_{22}} p_l & \frac{1}{n} &= \lambda_1 p_j + \lambda_2 p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k \\
\frac{1}{n} &= \lambda_1 p_k + \lambda_2 p_k \sum_{j:(X_j,Y_j)\in A_{12}} p_j & \frac{1}{n} &= \lambda_1 p_l - \lambda_2 p_l \sum_{i:(X_i,Y_i)\in A_{11}} p_i.
\end{aligned}$$

Therefore,

$$\begin{aligned}
P_n(A_{11}) &= \lambda_1 \sum_{i\in A_{11}} p_i - \lambda_2 \sum_{i\in A_{11}} p_i \sum_{l:(X_l,Y_l)\in A_{22}} p_l = \lambda_1 \tilde{P}(A_{11}) - \lambda_2 \tilde{P}(A_{11})\tilde{P}(A_{22}) \\
P_n(A_{12}) &= \lambda_1 \sum_{i\in A_{12}} p_j + \lambda_2 \sum_{j\in A_{12}} p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k = \lambda_1 \tilde{P}(A_{12}) + \lambda_2 \tilde{P}(A_{12})\tilde{P}(A_{21}) \\
P_n(A_{21}) &= \lambda_1 \sum_{k\in A_{21}} p_k + \lambda_2 \sum_{k\in A_{21}} p_k \sum_{j:(X_j,Y_j)\in A_{12}} p_j = \lambda_1 \tilde{P}(A_{21}) + \lambda_2 \tilde{P}(A_{21})\tilde{P}(A_{12}) \\
P_n(A_{22}) &= \lambda_1 \sum_{l\in A_{22}} p_l - \lambda_2 \sum_{l:(X_l,Y_l)\in A_{22}} p_l \sum_{i:(X_i,Y_i)\in A_{11}} p_i = \lambda_1 \tilde{P}(A_{22}) - \lambda_2 \tilde{P}(A_{22})\tilde{P}(A_{11}).
\end{aligned}$$

Note that $1 = \tilde{P}(A_{11})+\tilde{P}(A_{12})+\tilde{P}(A_{21})+\tilde{P}(A_{22})$ and $\tilde{P}(A_{11})\tilde{P}(A_{22}) = \tilde{P}(A_{12})\tilde{P}(A_{21})$ under the independence restriction. Summing all four empirical probability, we obtain

$$1 = \lambda_1 + \lambda_2 \times 0,$$

so that $\lambda_1 = 1$. Thus,

$$\begin{aligned}
P_n(A_{11}) &= \tilde{P}(A_{11}) - \lambda_2 \tilde{P}(A_{11})\tilde{P}(A_{22}) & P_n(A_{12}) &= \tilde{P}(A_{12}) + \lambda_2 \tilde{P}(A_{12})\tilde{P}(A_{21}) \\
P_n(A_{21}) &= \tilde{P}(A_{21}) + \lambda_2 \tilde{P}(A_{21})\tilde{P}(A_{12}) & P_n(A_{22}) &= \tilde{P}(A_{22}) - \lambda_2 \tilde{P}(A_{22})\tilde{P}(A_{11}).
\end{aligned}$$

Summing the first and second equations above, we obtain $F_n(x) = \tilde{F}_n(x)$. Summing the third and the fourth equations, we obtain $G_n(y) = \tilde{G}_n(y)$. According to the independence restriction, we obtain

$$\tilde{P}(A_{11}) = F_n(x)G_n(y) \qquad \tilde{P}(A_{12}) = F_n(x)\{1 - G_n(y)\}$$

$$\tilde{P}(A_{21}) = \{1 - F_n(x)\}G_n(y) \qquad \tilde{P}(A_{22}) = \{1 - F_n(x)\}\{1 - G_n(y)\}$$

and

$$\lambda_2 = \frac{\{H_n(x,y) - F_n(x)G_n(y)\}\{1 - F_n(x)\}\{1 - G_n(y)\}}{F_n(x)\{1 - F_n(x)\}G_n(y)\{1 - G_n(y)\}}.$$

Now we obtain the probability that should be assigned at the data points.

$$\begin{aligned}
\frac{1}{n} &= p_i + \lambda_2 p_i \sum_{i \in A_{22}} p_l = p_i \left[1 + \frac{\{H_n(x,y) - F_n(x)G_n(y)\}\{1 - F_n(x)\}\{1 - G_n(y)\}}{F_n(x)\{1 - F_n(x)\}G_n(y)\{1 - G_n(y)\}}\right] \\
&= p_i \left[1 + \frac{\{H_n(x,y) - F_n(x)G_n(y)\}}{F_n(x)G_n(y)}\right] = p_i \left(\frac{H_n(x,y)}{F_n(x)G_n(y)}\right) \\
&= p_i \left(\frac{P_n(A_{11})}{F_n(x)G_n(y)}\right) \\
\frac{1}{n} &= p_j - \lambda_2 p_j \sum_{k:(X_k,Y_k) \in A_{21}} p_k = p_j \left[1 + \frac{\{H_n(x,y) - F_n(x)G_n(y)\}\{1 - F_n(x)\}G_n(y)}{F_n(x)\{1 - F_n(x)\}G_n(y)\{1 - G_n(y)\}}\right] \\
&= p_j \left[1 - \frac{\{H_n(x,y) - F_n(x)G_n(y)\}}{F_n(x)\{1 - G_n(y)\}}\right] = p_j \left[\frac{\{G_n(y) - H_n(x,y)\}}{F_n(x)\{1 - G_n(y)\}}\right] \\
&= p_j \left[\frac{P_n(A_{12})}{F_n(x)\{1 - G_n(y)\}}\right] \\
\frac{1}{n} &= p_k - \lambda_2 p_k \sum_{j:(X_j,Y_j) \in A_{12}} p_j = p_k \left[1 - \frac{\{H_n(x,y) - F_n(x)G_n(y)\}F_n(x)\{1 - G_n(y)\}}{F_n(x)\{1 - F_n(x)\}G_n(y)\{1 - G_n(y)\}}\right] \\
&= p_k \left[1 - \frac{\{H_n(x,y) - F_n(x)G_n(y)\}}{\{1 - F_n(x)\}G_n(y)}\right] = p_k \left[\frac{\{F_n(x) - H_n(x,y)\}}{\{1 - F_n(x)\}G_n(y)}\right] \\
&= p_k \left[\frac{P_n(A_{21})}{F_n(x)\{1 - G_n(y)\}}\right] \\
\frac{1}{n} &= p_l + \lambda_2 p_l \sum_{i:(X_i,Y_i) \in A_{11}} p_i = p_l \left[1 + \frac{\{H_n(x,y) - F_n(x)G_n(y)\}F_n(x)G_n(y)}{F_n(x)\{1 - F_n(x)\}G_n(y)\{1 - G_n(y)\}}\right] \\
&= p_l \left[1 + \frac{\{H_n(x,y) - F_n(x)G_n(y)\}}{\{1 - F_n(x)\}\{1 - G_n(y)\}}\right] \\
&= p_l \left[\frac{\{P_n(A_{22})\}}{\{1 - F_n(x)\}\{1 - G_n(y)\}}\right].
\end{aligned}$$

Substituting these into the empirical likelihood function L, we have

$$R_{01}(x,y) = \left(\frac{\{F_n(x)\}\{G_n(y)\}}{P_n(A_{11})}\right)^{nP_n(A_{11})} \left(\frac{\{F_n(x)\}\{1-G_n(y)\}}{P_n(A_{12})}\right)^{nP_n(A_{12})}$$
$$\times \left(\frac{\{1-F_n(x)\{G_n(y)\}}{P_n(A_{21})}\right)^{nP_n(A_{21})} \left(\frac{\{1-F_n(x)\}\{1-G_n(y)\}}{P_n(A_{22})}\right)^{nP_n(A_{22})}.$$

B.3 Supremum of empirical likelihoods under PQD restriction

For $\sup\{L(\tilde{H}) : H(x,y) \geq F(x)G(y)\}$, since

$$a > (a+b)(a+c) \iff ad > bc,$$

for all $0 \leq a, b, c, d \leq 1$ with $a+b+c+d = 1$, it suffices to consider

$$\begin{aligned}
\mathcal{L}(\mathbf{p}) = \frac{1}{n}&\left\{\sum_{i:(X_i,Y_i)\in A_{11}} \ln p_i + \sum_{j:(X_j,Y_j)\in A_{12}} \ln p_j + \sum_{k:(X_k,Y_k)\in A_{21}} \ln p_k + \sum_{l:(X_l,Y_l)\in A_{22}} \ln p_l\right\} \\
&+ \lambda_1\left\{\sum_{i:(X_i,Y_i)\in A_{11}} p_i + \sum_{j:(X_j,Y_j)\in A_{12}} p_j + \sum_{k:(X_k,Y_k)\in A_{21}} p_k + \sum_{l:(X_l,Y_l)\in A_{22}} p_l - 1\right\} \\
&+ \lambda_2\left\{\sum_{j:(X_j,Y_j)\in A_{12}} p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k - \sum_{i:(X_i,Y_i)\in A_{11}} p_i \sum_{l:(X_l,Y_l)\in A_{22}} p_l\right\}.
\end{aligned}$$

We consider two cases: $H_n(x,y) \geq F_n(x)G_n(y)$ and $H_n(x,y) < F_n(x)G_n(y)$.

Case 1. $H_n(x,y) \geq F_n(x)G_n(y)$

Consider $H_n(x,y) \geq F_n(x)G_n(y)$. If $\lambda_2 = 0$, we obtain the partial derivative

$$\begin{aligned}
\frac{\partial\mathcal{L}(\mathbf{p})}{\partial p_i} &= -\frac{1}{np_i} + \lambda_1 & \frac{\partial\mathcal{L}(\mathbf{p})}{\partial p_j} &= -\frac{1}{np_j} + \lambda_1 \\
\frac{\partial\mathcal{L}(\mathbf{p})}{\partial p_k} &= -\frac{1}{np_k} + \lambda_1 & \frac{\partial\mathcal{L}(\mathbf{p})}{\partial p_l} &= -\frac{1}{np_l} + \lambda_1.
\end{aligned}$$

Setting these partial derivatives equal to zero, we find that

$$\begin{aligned}
\frac{1}{n} &= \lambda_1 p_i & \frac{1}{n} &= \lambda_1 p_j \\
\frac{1}{n} &= \lambda_1 p_k & \frac{1}{n} &= \lambda_1 p_l.
\end{aligned}$$

Summing all the derivatives at all data points, we obtain $\lambda_1 = 1$. Furthermore, $p_i = p_j = p_k = p_k = 1/n$ is feasible because $P_n(A_{11})P_n(A_{22}) \geq P_n(A_{12})P_n(A_{21})$ which is equivalent to $H_n(x,y) \geq F_n(x)G_n(y)$. If $\lambda_2 > 0$, then the following condition holds

$$\sum_{j:(X_j,Y_j)\in A_{12}} p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k - \sum_{i:(X_i,Y_i)\in A_{11}} p_i \sum_{l:(X_l,Y_l)\in A_{22}} p_l = 0. \tag{B.1}$$

Taking the partial derivative of (B.1), we have

$$\begin{aligned}
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_i} &= -\frac{1}{np_i} + \lambda_1 + \lambda_2 \sum_{l:(X_l,Y_l)\in A_{22}} p_l \\
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_j} &= -\frac{1}{np_j} + \lambda_1 - \lambda_2 \sum_{k:(X_k,Y_k)\in A_{21}} p_k \\
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_k} &= -\frac{1}{np_k} + \lambda_1 - \lambda_2 \sum_{j:(X_j,Y_j)\in A_{12}} p_j \\
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_l} &= -\frac{1}{np_l} + \lambda_1 + \lambda_2 \sum_{i:(X_i,Y_i)\in A_{11}} p_i.
\end{aligned}$$

Setting these partial derivatives equal to zero and solving, we obtain

$$\begin{aligned}
\frac{1}{n} &= \lambda_1 p_i + \lambda_2 p_i \sum_{i\in A_{22}} p_l & \qquad \frac{1}{n} &= \lambda_1 p_j + \lambda_2 p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k \\
\frac{1}{n} &= \lambda_1 p_k + \lambda_2 p_k \sum_{j:(X_j,Y_j)\in A_{12}} p_j & \qquad \frac{1}{n} &= \lambda_1 p_l + \lambda_2 p_l \sum_{i:(X_i,Y_i)\in A_{11}} p_i.
\end{aligned}$$

So the empirical probability measure can be written as

$$\begin{aligned}
P_n(A_{11}) &= \lambda_1 \sum_{i\in A_{11}} p_i + \lambda_2 \sum_{i\in A_{11}} p_i \sum_{l:(X_l,Y_l)\in A_{22}} p_l = \lambda_1 \tilde{P}(A_{11}) + \lambda_2 \tilde{P}(A_{11})\tilde{P}(A_{22}) \\
P_n(A_{12}) &= \lambda_1 \sum_{i\in A_{12}} p_j - \lambda_2 \sum_{j\in A_{12}} p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k = \lambda_1 \tilde{P}(A_{12}) - \lambda_2 \tilde{P}(A_{12})\tilde{P}(A_{21}) \\
P_n(A_{21}) &= \lambda_1 \sum_{k\in A_{21}} p_k - \lambda_2 \sum_{k\in A_{21}} p_k \sum_{j:(X_j,Y_j)\in A_{12}} p_j = \lambda_1 \tilde{P}(A_{21}) - \lambda_2 \tilde{P}(A_{21})\tilde{P}(A_{12}) \\
P_n(A_{22}) &= \lambda_1 \sum_{l\in A_{22}} p_l + \lambda_2 \sum_{l:(X_l,Y_l)\in A_{22}} p_l \sum_{i:(X_i,Y_i)\in A_{11}} p_i = \lambda_1 \tilde{P}(A_{22}) + \lambda_2 \tilde{P}(A_{22})\tilde{P}(A_{11}).
\end{aligned}$$

We sum all the equations above and obtain $1 = \lambda_1 + \lambda_2 \times 0$ so that $\lambda_1 = 1$. Thus,

$$P_n(A_{11}) = \tilde{P}(A_{11}) + \lambda_2 \tilde{P}(A_{11})\tilde{P}(A_{22})$$

$$P_n(A_{12}) = \tilde{P}(A_{12}) - \lambda_2 \tilde{P}(A_{12})\tilde{P}(A_{21})$$

$$P_n(A_{21}) = \tilde{P}(A_{21}) - \lambda_2 \tilde{P}(A_{21})\tilde{P}(A_{12})$$

$$P_n(A_{22}) = \tilde{P}(A_{22}) + \lambda_2 \tilde{P}(A_{22})\tilde{P}(A_{11}).$$

Further, we have $\tilde{P}(A_{11}) = F_n(x)G_n(x)$, $\tilde{P}(A_{12}) = F_n(x)\{1 - G_n(x)\}$, $\tilde{P}(A_{21}) = \{1 - F_n(x)\}G_n(x)$, $\tilde{P}(A_{22}) = \{1 - F_n(x)\}\{1 - G_n(x)\}$, and

$$\lambda_2 = \frac{\{H_n(x,y) - F_n(x)G_n(y)\}}{F_n(x)\{1 - F_n(x)\}G_n(y)\{1 - G_n(y)\}}.$$

However, $\tilde{P}(A_{11})$, $\tilde{P}(A_{12})$, $\tilde{P}(A_{21})$, $\tilde{P}(A_{22})$ do not satisfy the condition in Equation (B.1). So the only solution is $p_i = p_j = p_k = p_l = 1/n$.

Case 2. $H_n(x,y) < F_n(x)G_n(y)$

If $H_n(x,y) < F_n(x)G_n(y)$ and if $\lambda_2 = 0$, then $\lambda_1 = 1$ but $p_i = p_j = p_k = p_k = 1/n$ is not feasible because $P_n(A_{11})P_n(A_{22}) < P_n(A_{12})P_n(A_{21})$. If $\lambda_2 > 0$, then the condition in Equation (B.1) must hold. Taking the partial derivatives of Equation (B.1), we obtain

$$\begin{aligned}
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_i} &= -\frac{1}{np_i} + \lambda_1 + \lambda_2 \sum_{l:(X_l,Y_l)\in A_{22}} p_l \\
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_j} &= -\frac{1}{np_j} + \lambda_1 - \lambda_2 \sum_{k:(X_k,Y_k)\in A_{21}} p_k \\
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_k} &= -\frac{1}{np_k} + \lambda_1 - \lambda_2 \sum_{j:(X_j,Y_j)\in A_{12}} p_j \\
\frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_l} &= -\frac{1}{np_l} + \lambda_1 + \lambda_2 \sum_{i:(X_i,Y_i)\in A_{11}} p_i.
\end{aligned}$$

Setting the partial derivatives above equal to zero and solving, we obtain

$$\begin{aligned}
\frac{1}{n} &= \lambda_1 p_i + \lambda_2 p_i \sum_{i\in A_{22}} p_l & \frac{1}{n} &= \lambda_1 p_j - \lambda_2 p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k \\
\frac{1}{n} &= \lambda_1 p_k - \lambda_2 p_k \sum_{j:(X_j,Y_j)\in A_{12}} p_j & \frac{1}{n} &= \lambda_1 p_l + \lambda_2 p_l \sum_{i:(X_i,Y_i)\in A_{11}} p_i.
\end{aligned}$$

Therefore,

$$P_n(A_{11}) = \lambda_1 \sum_{i \in A_{11}} p_i + \lambda_2 \sum_{i \in A_{11}} p_i \sum_{l:(X_l,Y_l)\in A_{22}} p_l = \lambda_1 \tilde{P}(A_{11}) + \lambda_2 \tilde{P}(A_{11})\tilde{P}(A_{22})$$

$$P_n(A_{12}) = \lambda_1 \sum_{i \in A_{12}} p_j - \lambda_2 \sum_{j \in A_{12}} p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k = \lambda_1 \tilde{P}(A_{12}) - \lambda_2 \tilde{P}(A_{12})\tilde{P}(A_{21})$$

$$P_n(A_{21}) = \lambda_1 \sum_{k \in A_{21}} p_k - \lambda_2 \sum_{k \in A_{21}} p_k \sum_{j:(X_j,Y_j)\in A_{12}} p_j = \lambda_1 \tilde{P}(A_{21}) - \lambda_2 \tilde{P}(A_{21})\tilde{P}(A_{12})$$

$$P_n(A_{22}) = \lambda_1 \sum_{l \in A_{22}} p_l + \lambda_2 \sum_{l:(X_l,Y_l)\in A_{22}} p_l \sum_{i:(X_i,Y_i)\in A_{11}} p_i = \lambda_1 \tilde{P}(A_{22}) + \lambda_2 \tilde{P}(A_{22})\tilde{P}(A_{11}).$$

Summing all the equations above, we obtain $1 = \lambda_1 + \lambda_2 \times 0$, i.e., $\lambda_1 = 1$ so that

$$P_n(A_{11}) = \tilde{P}(A_{11}) + \lambda_2 \tilde{P}(A_{11})\tilde{P}(A_{22}) \quad P_n(A_{12}) = \tilde{P}(A_{12}) - \lambda_2 \tilde{P}(A_{12})\tilde{P}(A_{21})$$

$$P_n(A_{21}) = \tilde{P}(A_{21}) - \lambda_2 \tilde{P}(A_{21})\tilde{P}(A_{12}) \quad P_n(A_{22}) = \tilde{P}(A_{22}) + \lambda_2 \tilde{P}(A_{22})\tilde{P}(A_{11}).$$

Then we have $F_n(x) = \tilde{F}_n(x)$ and $G_n(y) = \tilde{G}_n(y)$. Further, we have $\tilde{P}(A_{11}) = F_n(x)G_n(y)$, $\tilde{P}(A_{12}) = F_n(x)\{1 - G_n(y)\}$, $\tilde{P}(A_{21}) = \{1 - F_n(x)\}G_n(y)$, $\tilde{P}(A_{22}) = \{1 - F_n(x)\}\{1 - G_n(y)\}$, and

$$\lambda_2 = \frac{\{H_n(x,y) - F_n(x)G_n(y)\}}{F_n(x)\{1 - F_n(x)\}G_n(y)\{1 - G_n(y)\}}.$$

In this case,

$$\begin{aligned}
\frac{1}{n} &= p_i + \lambda_2 p_i \sum_{i \in A_{22}} p_l = p_i \left[1 + \frac{\{H_n(x,y) - F_n(x)G_n(y)\}\{1 - F_n(x)\}\{1 - G_n(y)\}}{F_n(x)\{1 - F_n(x)\}G_n(y)\{1 - G_n(y)\}}\right] \\
&= p_i \left[1 + \frac{H_n(x,y) - F_n(x)G_n(y)}{F_n(x)G_n(y)}\right] = p_i \left(\frac{H_n(x,y)}{F_n(x)G_n(y)}\right) = p_i \left(\frac{P_n(A_{11})}{F_n(x)G_n(y)}\right) \\
\frac{1}{n} &= p_j - \lambda_2 p_j \sum_{k:(X_k,Y_k)\in A_{21}} p_k = p_j \left[1 + \frac{\{H_n(x,y) - F_n(x)G_n(y)\}\{1 - F_n(x)\}G_n(y)}{F_n(x)\{1 - F_n(x)\}G_n(y)\{1 - G_n(y)\}}\right] \\
&= p_j \left[1 - \frac{H_n(x,y) - F_n(x)G_n(y)}{F_n(x)\{1 - G_n(y)\}}\right] = p_j \left[\frac{G_n(y) - H_n(x,y)}{F_n(x)\{1 - G_n(y)\}}\right] \\
&= p_j \left[\frac{P_n(A_{12})}{F_n(x)\{1 - G_n(y)\}}\right]
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{n} &= p_k - \lambda_2 p_k \sum_{j:(X_j,Y_j)\in A_{12}} p_j = p_k \left[1 - \frac{\{H_n(x,y) - F_n(x)G_n(y)\}F_n(x)\{1-G_n(y)\}}{F_n(x)\{1-F_n(x)\}G_n(y)\{1-G_n(y)\}}\right] \\
&= p_k \left[1 - \frac{H_n(x,y) - F_n(x)G_n(y)}{\{1-F_n(x)\}G_n(y)}\right] = p_k \left[\frac{F_n(x) - H_n(x,y)}{\{1-F_n(x)\}G_n(y)}\right] \\
&= p_k \left[\frac{P_n(A_{21})}{F_n(x)\{1-G_n(y)\}}\right] \\
\frac{1}{n} &= p_l + \lambda_2 p_l \sum_{i:(X_i,Y_i)\in A_{11}} p_i = p_l \left[1 + \frac{\{H_n(x,y) - F_n(x)G_n(y)\}F_n(x)G_n(y)}{F_n(x)\{1-F_n(x)\}G_n(y)\{1-G_n(y)\}}\right] \\
&= p_l \left[1 + \frac{H_n(x,y) - F_n(x)G_n(y)}{\{1-F_n(x)\}\{1-G_n(y)\}}\right] = \ p_l \left[\frac{P_n(A_{22})}{\{1-F_n(x)\}\{1-G_n(y)\}}\right].
\end{aligned}
$$

Substituting these into the empirical likelihood function L, we have $R(x,y) = 1$ if $P_n(A_{11}) \geq F_n(x)G_n(y)$. If $P_n(A_{11}) < F_n(x)G_n(y)$, we have

$$
\begin{aligned}
R_{12}(x,y) = \ &\left(\frac{F_n(x)G_n(y)}{P_n(A_{11})}\right)^{nP_n(A_{11})} \left(\frac{F_n(x)\{1-G_n(y)\}}{P_n(A_{12})}\right)^{nP_n(A_{12})} \\
&\times \left(\frac{\{1-F_n(x)\}G_n(y)}{P_n(A_{21})}\right)^{nP_n(A_{21})} \left(\frac{\{1-F_n(x)\}\{1-G_n(y)\}}{P_n(A_{22})}\right)^{nP_n(A_{22})}.
\end{aligned}
$$

B.4 Critical values

Here we list the critical values that we used in simulations in Chapter 3. Table B.1 lists critical values for testing $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$. Table B.2 lists critical values for testing $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$.

Table B.1: Estimated critical values for test statistics generated by the independence copula with significance level $\alpha = 0.1, 0.05, 0.025, 0.01$. $10,000$ Monte Carlo samples of size $n = 10, 30, 50, 100, 200$ are used to estimate the critical values.

$n = 10$	90%	95%	97.5%	99%
EL_1	1.988	2.501	2.977	3.600
EL_2	1.133	1.431	1.721	2.027
EM	1.506	1.756	1.991	2.277
KS_1	0.643	0.666	0.750	0.800
CvM_1	0.162	0.195	0.225	0.263
AD_1	6.596	7.845	8.929	10.100
$n = 30$	90%	95%	97.5%	99%
EL_1	1.667	2.101	2.497	2.963
EL_2	1.140	1.444	1.751	2.106
EM	1.564	1.850	2.114	2.449
KS_1	0.611	0.675	0.732	0.800
CvM_1	0.089	0.109	0.130	0.159
AD_1	4.584	5.637	6.532	7.733
$n = 50$	90%	95%	97.5%	99%
EL_1	1.556	1.926	2.313	2.827
EL_2	1.117	1.400	1.728	2.085
EM	1.583	1.877	2.149	2.584
KS_1	0.610	0.670	0.727	0.790
CvM_1	0.072	0.092	0.109	0.134
AD_1	3.828	4.759	5.613	6.693
$n = 100$	90%	95%	97.5%	99%
EL_1	1.369	1.732	2.086	2.526
EL_2	1.076	1.364	1.711	2.097
EM	1.536	1.816	2.096	2.472
KS_1	0.604	0.667	0.727	0.798
CvM_1	0.056	0.071	0.089	0.108
AD_1	2.990	3.671	4.421	5.500
$n = 200$	90%	95%	97.5%	99%
EL_1	1.291	1.626	1.974	2.437
EL_2	1.080	1.370	1.711	2.123
EM	1.538	1.816	2.118	2.535
KS_1	0.618	0.684	0.741	0.808
CvM_1	0.049	0.063	0.078	0.098
AD_1	2.392	2.911	3.450	4.352

Table B.2: Estimated critical values for test statistics generated by the independence copula with significance level $\alpha = 0.1, 0.05, 0.025, 0.01$. $10,000$ Monte Carlo samples of size $n = 10, 30, 50, 100, 200$ are used to estimate the critical values.

$n = 10$	90%	95%	97.5%	99%
EL_3	1.122	1.412	1.675	1.997
EL_4	0.592	0.827	1.078	1.466
KS_2	0.152	0.206	0.282	0.311
CvM_2	0.003	0.006	0.010	0.016
AD_2	0.056	0.128	0.197	0.337
$n = 30$	90%	95%	97.5%	99%
EL_3	1.124	1.417	1.717	2.112
EL_4	0.742	0.989	1.238	1.593
KS_2	0.329	0.387	0.454	0.512
CvM_2	0.011	0.017	0.023	0.034
AD_2	0.269	0.402	0.566	0.779
$n = 50$	90%	95%	97.5%	99%
EL_3	1.125	1.446	1.780	2.223
EL_4	0.782	1.076	1.373	1.797
KS_2	0.391	0.462	0.516	0.587
CvM_2	0.014	0.022	0.031	0.043
AD_2	0.376	0.568	0.775	1.058
$n = 100$	90%	95%	97.5%	99%
EL_3	1.101	1.386	1.672	2.098
EL_4	0.840	1.104	1.362	1.743
KS_2	0.458	0.520	0.578	0.647
CvM_2	0.019	0.027	0.036	0.048
AD_2	0.515	0.717	0.915	1.208
$n = 200$	90%	95%	97.5%	99%
EL_3	1.061	1.368	1.683	2.112
EL_4	0.883	1.152	1.422	1.838
KS_2	0.501	0.575	0.634	0.701
CvM_2	0.022	0.031	0.040	0.054
AD_2	0.609	0.853	1.072	1.417

Appendix C

Supplementary materials for Chapter 4

C.1 Supremum of empirical likelihoods under exchangeability restriction

To find the supremum of the empirical likelihood under an exchangeability assumption; i.e., $\sup\{L(\tilde{H}) : H(x,y) = H(y,x)\}$, we define

$$\begin{aligned}\mathcal{L}(\mathbf{p}) = \frac{1}{n}&\left\{\sum_{i:(X_i,Y_i)\in A_h} \ln p_i + \sum_{j:(X_j,Y_j)\in A_v} \ln p_j\right\}\\ &+ \lambda_1\left\{\sum_{i:(X_i,Y_i)\in A_h} p_i + \sum_{j:(X_j,Y_j)\in A_v} p_j + \sum_{k:(X_k,Y_k)\in A_r} p_k - 1\right\}\\ &+ \lambda_2\left\{\sum_{i:(X_i,Y_i)\in A_h} p_i - \sum_{j:(X_j,Y_j)\in A_v} p_j\right\}.\end{aligned}$$

To maximize $\mathcal{L}(\mathbf{p})$, take the partial derivative respect to p_i for i such that $(X_i, Y_i) \in A_h$ and p_j for j such that $(X_j, Y_j) \in A_v$. Then we obtain

$$\frac{\partial\mathcal{L}(\mathbf{p})}{\partial p_i} = -\frac{1}{np_i} + \lambda_1 + \lambda_2 \quad \frac{\partial\mathcal{L}(\mathbf{p})}{\partial p_j} = -\frac{1}{np_j} + \lambda_1 - \lambda_2 \quad \frac{\partial\mathcal{L}(\mathbf{p})}{\partial p_k} = -\frac{1}{np_k}.$$

Setting all partial derivatives equal to zero and solving, we obtain

$$\frac{1}{n} = \lambda_1 p_i - \lambda_2 p_i \quad \frac{1}{n} = \lambda_1 p_j + \lambda_2 p_j \quad \frac{1}{n} = \lambda_1 p_k.$$

Therefore,

$$P_n(A_h) = (\lambda_1 - \lambda_2)\sum_{i\in A_h} p_i \quad P_n(A_v) = (\lambda_1 + \lambda_2)\sum_{i\in A_v} p_j \quad P_n(A_r) = \lambda_1 \sum_{k\in A_r} p_k.$$

Note that $1 = \tilde{P}(A_h) + \tilde{P}(A_v) + \tilde{P}(A_r)$ and $\tilde{P}(A_h) = \tilde{P}(A_v)$ under the exchangeability restriction. Summing all three empirical probabilities, we obtain $1 = \lambda_1 + \lambda_2 \times 0$, so that $\lambda_1 = 1$. Thus, $P_n(A_h) = (1 - \lambda_2)\tilde{P}(A_h)$, $P_n(A_v) = (1 + \lambda_2)\tilde{P}(A_v)$, $P_n(A_r) = \tilde{P}(A_r)$, and

$$\lambda_2 = \frac{P_n(A_h) - P_n(A_v)}{\tilde{P}(A_h) + \tilde{P}(A_v)} = \frac{P_n(A_h) - P_n(A_v)}{P_n(A_h) + P_n(A_v)}.$$

So we obtain the probability that should be assigned at the data points:

$$p_i = \frac{P_n(A_h) + P_n(A_v)}{2nP_n(A_h)}; \quad p_j = \frac{P_n(A_h) + P_n(A_v)}{2nP_n(A_v)}.$$

Substituting these into the empirical likelihood function L, we have

$$R(x, y) = \left(\frac{P_n(A_h) + P_n(A_v)}{2P_n(A_h)}\right)^{nP_n(A_h)} \left(\frac{P_n(A_h) + P_n(A_v)}{2P_n(A_v)}\right)^{nP_n(A_v)}.$$